Building Word-Problem Solving and Working Memory Capacity: A Randomized Controlled Trial Comparing Three Intervention Approaches

Lynn S. Fuchs[1,2], Douglas Fuchs[1,2], Sonya K. Sterba[1],

Marcia A. Barnes[1], Pamela M. Seethaler[1], and Paul Changas[3]

[1]Vanderbilt University, [2]American Institutes for Research,

[3]Metropolitan-Nashville Public Schools

Inquiries should be sent to Lynn S. Fuchs, lynn.fuchs@vanderbilt.edu; 228 Peabody, Vanderbilt University, Nashville, TN 37203.

This research was supported by R305A150200 and Core Grant HD15052 to Vanderbilt University from the Institute of Education Science's National Center on Education Research in the U.S. Department of Education. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Institute of Education Science or the National Center on Education Research or the U.S. Department of Education.

Fuchs, L.S., Fuchs, D., Sterba, S.K., Barnes, M.A., Seethaler, P.M., & Changas, P. (2022). Building word-problem solving and working memory capacity: A randomized controlled trial comparing three intervention approaches. *Journal of Educational Psychology*. Advance online publication. https://doi.org/10.1037/edu0000752

Date accepted: March 23, 2022

Abstract

This study's purpose was to investigate effects of 3 intervention approaches for building working memory (WM) and improving word-problem solving (WPS). Children with mathematics difficulties (n = 240; 7.51 years [*SD* = 0.33]) were randomized to 4 conditions: a control group, general WM training with contiguous math practice, WPS intervention without WM training, and WPS intervention with domain-specific WM training. WM, WPS, and arithmetic were assessed before and 1-3 weeks after intervention; delayed WPS and arithmetic posttesting occurred 4-6 weeks later. Multilevel modeling of main effects and mediation effects were employed. Compared to control, general WM training with contiguous math practice and WPS intervention without WM training increased WM and WPS. The 3rd training condition, WPS intervention with domain-specific WM training, which minimized WM training time, improved WPS but without effects on WM. Both WPS intervention conditions outperformed general WM training on WPS. Conclusions are as follows. (1) General WM training with contiguous math practice improves WM and WPS. (2) WM training is not a substitute for WPS intervention when the goal is to strengthen WPS. (3) WPS intervention without WM training improves WM but is not a substitute for WM training when the goal is to strengthen WM. (4) For WM effects to accrue, WM training needs to occur with sufficient intensity. (5) WM plays a causal role in WPS, but not in arithmetic. Implications are drawn for research and practice, including assessing instructional supports in future research to build cognitive-academic bidirectionality.

Key Words: working memory training, transfer, math problem solving, arithmetic

Educational Impact and Implications Statement

This study provides the basis for the following conclusions about interventions to improve working memory and word-problem solving for second graders with mathematics difficulties. First, general working memory training with math practice (without word-problem solving intervention) improves working memory as well as word-problem solving. Yet, working memory training is not a substitute for word-problem solving intervention when the goal is to strengthen word-problem solving, because WPS outcome is stronger with WPS intervention. Conversely, word-problem solving intervention (without working memory training) improves working memory. Yet, word-problem solving intervention is not a substitute for working memory training when the goal is to strengthen working memory, because WM outcome is stronger with general WM training.

Building Word-Problem Solving and Working Memory Capacity:

A Randomized Controlled Trial Comparing Three Intervention Approaches

Word-problem solving (WPS) is important. It reflects the capacity to apply mathematical ideas in everyday life and in science, technology, and engineering; it supports advanced mathematics learning (Hoffer et al., 2007); and it is a school-age predictor of employment and wages in adulthood (Batty et al., 2010; Every Child a Chance Trust 2009). Yet WPS difficulty is widespread (Daroczy et al., 2015) and can occur even when the calculation skill required for problem solution is intact (Cummins et al., 1988; Koedinger & Nathan, 2004). Differential difficulty with WPS may be due to its stronger reliance on a variety of cognitive resources than are engaged during calculations (Fuchs et al., 2018; Swanson & Beebe-Frankenberger, 2004). This includes working memory (WM), the storage and manipulation of a limited amount of information over a short amount of time (Cowan, 2014).

WM capacity features prominently in theoretical accounts of WPS. According to Kintsch and Greeno (1985), for example, propositions in the word-problem narrative trigger a series of set-building strategies that rely on WM to iteratively incorporate relevant ideas within a word-problem schema. Consider the following problem-solving process for this second-grade combine word problem (Part 1 plus Part 2 equals Total or P1 + P2 = T): *Cleo has 6 sisters. Her cousin, Flora, has 2 brothers and 3 sisters. How many sisters do the girls have in all?* A problem solver may process the first sentence's propositional text base to identify that the object is sisters, the quantity is 6, and the actor is Cleo, whose role is to be determined. The problem solver stores this information in WM. In the second sentence, the problem solver similarly codes the propositions and places them in WM but notes that *brothers* fails to match the object code in the first sentence, signaling 2 as possibly irrelevant. This possibility is added to WM. In the last sentence (the question), the problem solver is cued by the quantitative proposition *how many sisters* and the phrase *in all* to select the combine schema; assign the role of superset (Total) to the missing quantity; assign subset roles (Part 1 and Part 2) to the to-be-determined items in WM; and reject 2 *brothers* as irrelevant. The problem solver translates the numbers associated with the combine schema's slots to build a number sentence with a missing quantity and calculates the solution.

In line with such accounts, research provides support for the importance of WM in WPS. Students with stronger versus weaker WPS differ on WM (Passolunghi & Siegel, 2004; Swanson & Sachse-Lee, 2001); WM has been shown to moderate the effects of WPS interventions (Swanson, 2016; Swanson et al., 2014); and individual differences in WM account for variance in WPS when controlling for other cognitive resources (Fuchs et al., 2010, 2020; Swanson & Beebe-Frankenberger, 2004). Peng et al. (2016) meta-analyzed the relation between WM and mathematics, while controlling for age and while testing WM modality, type of math skill, and type of students' math difficulty as moderators of this relation. In 100 studies with 829 estimates, the math skill with the largest correlation was WPS (r = .37), and this relation was comparable across verbal, numerical, and visuospatial WM modalities.

Given theories of WPS that rely on WM along with empirical evidence that WM helps explain individual differences in WPS, it seems plausible that strengthening WM can improve WPS. Yet, estimates of transfer from WM training to other forms of mathematics performance have been disappointing, and effects of WM training on WPS have not been investigated. In Melby-Lervag and Hulme's (2013) meta-analysis, WM training produced reliable short-term improvements in verbal WM, with a mean effect size (ES) of 0.56. Seven of the studies in that synthesis investigated transfer to math, all focused on calculations, with a mean ES of 0.25 (excluding two studies that targeted participants with IQs from 55 to 85). Sala and Gobet's 2017 meta-analysis, which was restricted to learners without learning difficulties, revealed a significant effect on WM (ES = 0.46), while an ES of 0.20 on math was moderated by study quality. As with the earlier meta-analysis, no study focused on WPS. (Zhang et al. [2018]

assessed *problem solving*, but did not define the term, and the study outcome was aggregated across calculations and problem solving.)

The purpose of the present study was to investigate the effects of alternative intervention approaches for building WM capacity and improving WPS. In this report, we focus on the study's primary outcome measures. The study's target population was students with mathematics difficulties (WPS performance at study entry below the 30th percentile), due to this population's challenges with WPS and WM. We focused on second grade given (a) cognitive malleability for building WM capacity at young ages (Zhang et al., 2019) and (b) the increasing complexity of second-grade WPS learning standards, which creates greater opportunity for WM engagement during WPS than at lower grades. In Peng et al. (2016), the correlation between WM and WPS for children of second-grade age was .45.

Although there are various theories of WM, this study's methods reflect current views of WM as a domain-general capacity, in which a unitary WM maintenance system retains stimuli from different modalities while retaining the information's modality-specific attributes. For example, in accounting for variability in verbal and visual WM measures, Kane et al.'s (2004) best-fitting latent model required a general WM factor. In exploring brain networks involved in visual and auditory WM (Li et al., 2014), domain-specific networks were involved during stimulus encoding but not in other periods of WM; by contrast, domain-general networks were sensitive to WM load across encoding, maintenance, and retrieval. Accordingly, the present study's WM outcome involved a latent WM variable with multiple indicators: a visuospatial complex WM task, a verbal complex WM task involving sentences, and a verbal complex WM task involving numerical stimuli.

At the same time, because domain-specific networks appear involved during stimulus encoding and because WM retains information-specific modalities (Li et al., 2014), the present randomized controlled trial was focused in part on two WM training conditions designed to increase connections between processes engaged during WM training sessions and math, consistent with transfer-appropriate processing theory (Franks et al., 2000; Roediger et al., 1989). One condition involved general WM training with brief, contiguous math practice at the end of each session. The other combined WM training exercises conducted on math stimuli with WPS intervention. The third training condition involved the same WPS intervention but without WM training exercises. The fourth was a business-as-usual control group (the conventional school program).

This Study's WM Training Conditions: Facilitating Transfer to Mathematics

Why Transfer to Math Challenges Students with MD

Two problems may explain limited transfer from WM training to math in students with mathematics difficulty (MD). The first is that students with MD experience general difficulty with transfer, often failing to recognize novel stimuli as related to tasks on which they have received instruction (Haskell 2001; National Research Council 2000). Randomized controlled trials on WPS instruction show that increases in WP transfer distance (degree of alignment between instructional content and outcome measures) and have a more deleterious effect on the performance of students with MD than on students who are average or high achieving (Fuchs, Fuchs, Craddock, et al., 2008). For example, students with MD who are taught to successfully solve word problems in which two quantities are combined to form a total frequently fail to recognize word problems with three quantities to be combined as requiring a similar problem model and solution strategy (Powell et al., 2009). This suggests that, with WM training, students with MD may fail to recognize opportunities for applying their increasing WM capacity in the context of math. Therefore, transfer may be facilitated by providing math practice contiguously with WM training or by using math stimuli for WM training.

A second problem likely contributing to limited transfer from WM training to math is that students with MD enter WM training with poor math skill. In WPS, students with MD often process problems superficially by finding key words to select an operation or by adding known quantities without using the narrative to build a problem model (Fuchs et al., 2021). Without a foundation of sound WPS strategies, it is unlikely that increasing WM capacity will transfer to improved WPS outcomes. This suggests that WM training for students with MD needs to be conducted in conjunction with instruction designed to teach productive strategies for building problem models.

Previous Studies on WM Training Designed to Account for Limited Transfer in Students with MD

We identified three randomized controlled trials, all conducted since the Melby-Lervag and Hulme meta-analysis (2013), testing the effects of WM training protocols designed to account for one or the other explanation for limited transfer among students with MD. Note that in considering transfer from WM training to math, we include studies with more lenient definitions of MD than used in the present study. This was necessary because including a broader set of studies permitted us to consider studies of greatest relevance.

Nelwan and Kroesbergen (2016) focused on building mathematics skill. Students (9-12 years of age) with pretest math scores below the 50th percentile were randomized to three conditions, each with two 8-week periods: WM training followed by math training (adaptive computerized arithmetic training), math training followed by WM training, and non-active control followed by math training. Short-term effects of WM training occurred on verbal updating, but not on short- or long-term visuospatial updating and with little evidence of transfer to arithmetic fluency. Also, neither WM nor math training improved number sense. The authors speculated that disappointing results were due to inadequate supervision during training.

Nelwan et al. (2018) pursued this possibility with additional study participants randomly assigned to two conditions: (a) WM training then math training, with close monitoring of student progress and weekly feedback to address student difficulties, or (b) non-active control then math training. Analyses incorporated the 2016 study's students in the low-supervision WM training then math training students as a contrast condition. Results provided scant evidence that stronger supervision improves WM more than weaker supervision or control: Effects were significant only at the first of two outcome assessments and only on short-term visuospatial, not verbal WM. Also, WM training did not prime students in ways that differentially benefited them in the second phase's math training.

Kroesbergen et al. (2014) instead used numerical stimuli within WM training. This was contrasted to WM training that used similar exercises with domain-general stimuli and to a control condition. Five-year old children with math performance below the 50th percentile were randomized to the three conditions. Those in the two treatment conditions (general and domain-specific WM training) completed eight 30-min sessions over 4 weeks. There were no significant differences on phonological WM. On visuospatial WM, the two WM training conditions performed more strongly than control but comparably to each other. On nonsymbolic quantity discrimination, both WM training conditions outperformed the control group. On counting skills, domain-specific but not domain-general WM training outperformed control, but the two WM training conditions performed comparably.

A fourth study, although not involving students with MD, is of interest because WM training involved numerical stimuli. Rode et al. (2014) assigned third-grade classrooms to domain-specific WM training or conventional schooling. Training produced improvement on the WM training task, with evidence of transfer to measures of verbal and visuospatial WM. Transfer to math performance was, however, mixed: A significant effect was identified on one measure (ES = 0.26), without transfer on the other (ES = 0.05). Neither measure addressed WPS.

The pattern of results in Kroesbergen et al. (2014) and Rode et al. (2014), although mixed, suggests that processing domain-specific items for WM storage may support learning of new skills in that domain. This is consistent with transfer-appropriate processing theory where transfer is optimized when there is congruency between the processes used during learning and those required in the targeted outcome (Franks et al., 2000; Roediger et al., 1989). The pattern thus provides the basis for pursuing the idea that domain-specific WM training may enhance math outcomes. Note, however, that none of these prior studies focused on WPS. Also, these prior studies used a lenient criterion for MD, complicating conclusions for the population of learners for whom transfer is especially challenging.

The Present Study's Two Training Conditions Designed to Facilitate Transfer

In the present study, we sought to build on and extend this literature with two training conditions. The first involved a commercial, widely distributed general WM training program that addresses visuospatial and verbal WM modalities with minimal reliance on math stimuli. Its effects on WM have been demonstrated (Gray et al., 2012; Shinaver et al., 2014), but transfer to academic performance has been inconsistent (Bergman Nutley & Söderqvist, 2017; Roche & Johnson, 2014). To support transfer to WPS by addressing students' difficulty recognizing stimuli as pertinent for tasks on which they have received training, we included 5 min of math practice at the end of each session. Math performance feedback was limited to providing correct answers for errors (there was no structured math instruction). This condition extends the WM training literature focused on transfer to math among students with MD, because it relies on a form of general WM training with previously demonstrated effects on WM while facilitating transfer by providing math practice at the end of each WM training session. In this paper, we refer to this condition as general WM training plus contiguous math practice (GWM+P).

The second training condition was a researcher-developed form of WM training designed to address both sources of transfer difficulty. To help students recognize math as a context for applying their increasing WM capacity, we followed Kroesbergen et al. (2014) by using math stimuli for WM training but extended that study by including word-problem stimuli. We further extended Kroesbergen et al. by building a foundation of WPS skill on which students may apply increasing WM capacity. To do this, we combined domain-specific WM training with validated WPS intervention. This condition also extends Nelwan et al. (2016, 2018) by addressing both sources of transfer difficulty and by relying on a validated form of WPS intervention. We refer to this condition as WPS intervention with math WM training items (WPS+MWM).

The Present Study's WPS Intervention

The efficacy of the WPS intervention used within the WPS+MWM condition has been demonstrated in multiple randomized controlled trials (What Works Clearinghouse 2009, 2021). As a form of schema-building instruction, it teaches children to conceptualize word problems within word-problem types (schemas). Although the effects of this WPS intervention on WPS have been demonstrated, the effects of this WPS or other WPS interventions on WM capacity have not been tested. This is possible because engaging in complex academic tasks may improve WM (as discussed below). Thus, the present study extends this literature on WPS schema instruction by assessing WM effects. Hence, our third training arm was the same WPS intervention used in WPS+MWM, but without WM training. In this paper, we refer to this condition as WPS intervention.

Few studies have tested the idea that math intervention can improve WM, and results are mixed. For example, Nelwan and Kroesbergen (2016) and Nelwan et al. (2018) found no support for this idea with arithmetic training in 9-12 year old students, whereas Ramani et al. (2017) found that a 10-session counting-on numerical magnitude comparison game improved kindergarteners' WM. Given that WPS may account for more variance in WM than do other math skills (Peng et al., 2016), testing whether improving WPS via WPS intervention also benefits WM is important for two reasons: because WPS intervention may itself represent a form of WM training as students engage in WM updating to solve word problems (Kintsch & Greeno, 1985) and because some evidence indicates that high quality schooling triggers cognitive-academic bidirectionality (Peng & Kievit, 2020), a point we return to in the discussion.

Extensions Spanning the Three Training Conditions, Research Questions, and Hypotheses

Beyond the just-described extensions specific to each training condition, the present study adds to the WM training literature more broadly with its novel focus on WPS outcomes. Given Peng et al.'s (2016) meta-analytic finding, it is possible that WPS, a form of complex mathematical reasoning, may provide greater opportunity for transfer from WM training than is the case with the less complex math transfer targets addressed in prior work. To gain insight into this possibility, we also assessed arithmetic outcome, the dominant outcome in the WM training literature, and we explored whether posttest WM mediates significant intervention condition effects specifically on delayed-posttest WPS, not on delayed-posttest arithmetic. We hypothesized this is the case.

Our three main questions focused on the WPS outcome. First, does general WM training with contiguous math practice (GWM+P) improve WM and WPS performance and are significant effects of GWM+P on delayed-posttest WPS mediated via posttest WM performance? We hypothesized that GWM+P strengthens complex WM span at posttest, which improves WPS by permitting students to store and process information more efficiently during contiguously presented math practice. We thus hypothesized that strengthened posttest WM mediates the GWM+P intervention's effects on delayed-posttest WPS.

Second, does WPS intervention improve WM and WPS outcomes and are significant effects on delayed-posttest WPS mediated by posttest WM performance? We hypothesized that WPS intervention's extensive practice on the WM updating processes thought to be engaged during WPS (see account based on Kintsch & Greeno, 1985) represents a form of WM training. We thus hypothesized that WPS intervention strengthens complex WM span, which contributes to WPS intervention's effects on the delayed-posttest WPS outcome. We expected posttest WM performance to partially mediate WPS intervention's effects on delayed-posttest WPS.

Third, does combining WPS intervention with domain-specific WM training (WPS+MWM) improve WM and WPS outcomes and are significant effects on delayed-posttest WPS mediated by posttest WM performance? We expected WPS+MWM intervention to confer advantage on WPS outcomes over the each of the other training conditions. We expected stronger WPS outcomes at posttest and delayed posttest for the WPS+MWM intervention over the general WM training with math practice (GWM+P) for two reasons. First, consistent with transfer-appropriate processing theory (Franks et al., 2000; Roediger et al., 1989), we expected it to optimize congruency between processes used during the training sessions with those required in the targeted outcome. The second reason was that the WPS+MWM condition incorporated WPS intervention to build WPS strategies, onto which increased WM capacity may be applied. At the same time, we hypothesized stronger WPS outcome at posttest and delayed posttest for WPS+MWM intervention over WPS intervention without WM training, given expectations that domain-specific WM training may strengthen WM outcomes, which improves WPS outcomes by permitting students to store and process information more efficiently during WPS intervention's contiguously presented math practice. We expected posttest WM to mediate these treatment contrasts as well as the contrast between WPS+MWM intervention and the control group.

Before describing study methods, we note three methodological extensions to the WM training literature. First, we incorporated multilevel structural equation modeling to permit a focus on WM effects at the level of the underlying latent WM construct rather than manifest WM tasks. Second, we tested whether posttest WM mediates math effects assessed at delayed posttest (4-6 weeks after the measurement of the WM outcome). In the context of a randomized controlled trial, such timing strengthens causal inference about WM training's effects on math. Third, we included a follow-along sample of average- and high-performance classmates to serve as a comparison group for judging the severity of the pre- and posttest WM and math performance gaps of this study's sample of students with MD.

Method

Participants

We conducted this study in accord with our university-approved IRB protocol, which is charged with ensuring compliance with ethical and legal standards and offered participating schools professional development opportunities related to the study's intervention after the study's completion. To determine sample size for students with MD, we conducted power analysis using the Monte Carlo facility of M*plus* 7.11 (Muthén & Muthén, 1998-2013), following Muthén and Muthén (2002). The sample was drawn from a large, diverse, urban and suburban county-wide school district in the southeastern United States. We screened 1,702 consented children in 134 second-grade classrooms in 16 schools on *Story Problems* (Jordan & Hanich, 2020); 380 children met the study's benchmark for low WPS (<30th percentile). In line with study inclusion criteria, we excluded 31 students scoring above the 60th percentile on *Automated Working Memory Assessment* (AWMA) *Listening Recall* and *Counting Recall* (AWMA; Alloway, 2012) (to permit benefit from WM training, while including students with a range of WM scores to permit analysis of WM as a mediator) and 31 with standard scores below 80 on both subtests of the *Wechsler Abbreviated Intelligence Scale* (WASI; Wechsler, 2011) (because the study's WPS intervention was designed and validated to address the needs of students whose intellectual ability falls within the broadly average range). Prior to random assignment, 15 moved to non-participating schools, and we excluded two students identified as almost entirely non-English speakers (to avoid false positives) and 15 with scheduling difficulties. A final cohort of 28 children, equally distributed across study conditions, was not completed due to the premature end of the 2020-2021 school year due to COVID-19. As per the WWC 4.1 standards handbook (https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf), "losing sample members after random assignment because of acts of nature is not considered attrition when the loss is likely to affect intervention and control group members in the same manner" (p.11).

The analytic sample was 258 children randomly assigned at the individual level to the four study conditions: control (CON), general WM with contiguous math practice (GWM+P), WP intervention with math WM training items (WPS+MWM), and WP intervention without WM training (WPS intervention). During their participation, spanning one school year, 17 (6 CON, 2 GWM+P, 5 WPS+MWM, 4 WPS intervention) children moved out of the school district (beyond the study's reach), and one WPS child's behavior indicated the wish to withdraw from the study. Thus, the final sample comprised 240 students with MD.

See Table 1 for pretest WM and math means and standard deviations (*SD*s) by condition. In CON, GWM+P, WPS+MWM, and WPS intervention, respectively, there were 56%, 48%, 45%, and 42% males. Race and ethnicity were as follows: 25%, 35%, 28%, and 17% African American, 19%, 13%, 7%, and 17% white non-Hispanic, 47%, 49%, 58%, and 53% white Hispanic, and 9%, 3%, 7%, and 13% other. The respective percentages receiving English services was 49, 51, 58, and 57; special education services, 5, 2, 10, and 17; and subsidized school lunch (an indicator of economically disadvantaged households), 58, 57, 55, and 53.

To identify the follow-along sample of classmates, we randomly sampled 167 classmates who (a) scored *above* the risk WPS criterion for MD on the WPS screener, (b) scored *above* the 60th percentile on the two WM measures (to avoid students with WM limitations), (c) scored *at or above* 80 on both WASI subtests, and (d) were *not* almost entirely non-English speakers. Ten moved to schools outside the study's reach, for a final sample of 157 non-MD classmates. Classmates were assessed on all pre- and posttests completed by students with MD; we did not include them in delayed posttesting due to end-of-school-year logistical challenges.

See Table 1 for pretest WM and math scores for classmates, who outperformed students with MD on each study measure. Pretest performance gaps between students with MD and classmates, expressed as the difference between means divided by the classmate *SD*, are in Table 2. Classmates were 52% male, and 22% were African American, 33% white non-Hispanic, 33% white Hispanic, and 11% other. Twenty-two percent received English services; 39% received subsidized school lunch. Thus, compared to classmates, students with MD were disproportionately Hispanic and economically disadvantaged.

Screening Measures

Story Problems (Jordan & Hanich, 2000) comprises 14 word problems representing combine, compare, or change schema, with scenarios all involving pennies. Problems require addition and subtraction (sums and minuends to 12) for solution. The tester reads each item aloud; students have 30 s to construct an answer and can ask for re-reading(s) as needed. The score is the number of correct number numerical answers. Sample-based α was .88. Criterion validity with the Iowa Test of Basic Skills - Problem Solving is .71. *WASI* (Wechsler, 2011) is a 2-subtest measure of general cognitive ability, comprising *Vocabulary* and *Matrix Reasoning* subtests (reliability > .92). *Vocabulary* assesses expressive vocabulary, verbal knowledge, memory, learning ability, and crystallized and general intelligence. Students identify pictures and define words. *Matrix Reasoning* measures nonverbal fluid reasoning and general intelligence. Students complete matrices with missing pieces. Sample-based α was .82 and .80, respectively. *Working Memory Test Battery for Children* (WMTB-C; Pickering & Gathercole, 2001) - *Listening Recall and Counting Recall* are described below.

Working Memory Measures Used as Outcomes and Mediators

We assessed WM, the study's hypothesized mediator, at pre- and posttest using three WM complex span tasks (one visuospatial; two verbal). We used *Automated Working Memory Assessment* (AWMA) *Odd-One Out* (Alloway, 2012) and two subtests from the WMTB-C (Pickering & Gathercole, 2001): *Listening Recall and Counting Recall*. Each of the three WM tasks has six items at span levels from 1-6 to 1-9. Passing four items at a level moves a child to the next level. At each level, the number of items to be remembered increases by one. Failing three items terminates the subtest. The score is trials correct. For *Odd-One-Out*, children see three shapes, each in a box shown in a row and identify the odd-one-out; after making odd-one-out determinations for a series of rows, they recall the location of each odd-one-out shape in correct order by tapping the correct box. For *Listening Recall*, they determine if a sentence is true; after making true/false determinations for a series of sentences, they recall the last word of each sentence in correct order. For *Counting Recall*, they count a set of 4, 5, 6, or 7 dots on a card; after counting a series of cards, they recall the counts in correct order. Sample-based α was .71 - .73 for the three measures at pre- and posttest. We also measured performance on a fourth indicator of the latent WM factor: an experimental WM task comprising combine, compare, and change word problems. We excluded this primary measure from the present report because of its low and nonsignificant loading on the latent WM factor. For more information on this task and other results that led to its exclusion, contact the first or third author.

Mathematics Outcome Measures

We measured WPS and arithmetic at pre-, post-, and delayed posttest. We abbreviated delayed posttests given children's limited availability for testing near the end of the school year. *Arithmetic* includes four subtests, each with 25 problems selected from sample units drawn from state standards. *Addition 0 – 12* comprises addition problems with sums from 0 to 12; *Addition 5 - 18*, addition problems with sums from 5 to 18; *Subtraction 0 – 12*, subtraction problems with minuends from 1 to 12; *Subtraction 5 - 18*, subtraction problems with minuends from 5 to 18. For each subtest, students have 1 min to write answers. We used total number of correct answers across the four subtests. Sample-based α was .98. At delayed posttest, nine items included addition and subtraction with missing quantities in any slot of the number sentence. Sample-based α was .92. Criterion validity with Wide Range Achievement Test – Arithmetic is .78.

Second-Grade Word Problems includes 12 word problems selected from sample units drawn from state standards: combine, compare, and change problems, with and without irrelevant information, with one or two steps required for solution. The tester reads a word problem aloud; children follow along on paper, with up to 2 min to write a response before the tester reads the next word problem. Each is scored for correct math (1 point) and label (1 point) to reflect processing of the problem statement and understanding of the problem's theme and to transform numerical answers to meaningful problem solutions. At pretest, to shorten administration time, six problems were selected to represent the full range of difficulty. At delayed posttest, we used the same abbreviated test. Sample-based α was .78 at pretest; .85 at posttest; and .83 at delayed posttest. Criterion validity with the Iowa Test of Basic Skills - Problem Solving is .68.

Intervention

When describing interventions, we used the present tense because these interventions are ongoing (used in other work). When describing other information about study conditions, we used the past tense to communicate those procedures have been completed. See Table 3 for minutes spent in training segments by condition.

The three active intervention conditions shared three features. First, 45 30-min sessions were conducted one-to-one over 15 weeks outside the classroom in the child's school; absences and snow days were made up, usually within 1 school week. Second, children completed the same 5 min practice sheet at the end of each session. One side of the practice sheet includes nine addition and subtraction problems with missing information in any of three slots of the number sentence; the other side is a word problem. Children have 2 min to complete arithmetic problems and 2 min to complete the word problem. Third, because students with MD often display difficulty with attention, motivation, and perseverance through difficult tasks (e.g., Fuchs et al., 2013), a self-regulation system was included in each condition. It centers on four rules: use inside voices; stay in seat; follow directions; and try hard to answer problems correctly. Tutors set a timer to beep at unpredictable intervals and award a checkmark if the child is following all four rules when it beeps. Also, practice sheets have pre-designated bonus problems, revealed to children only after completion; they earn a bonus point for each correctly answered bonus problem. Tutors keep track of checkmarks. At each session's end, checkmarks are converted to stickers. When the sticker chart is full (~weekly), the child picks a small prize.

General Working Memory Training with Contiguous Math Practice (GWM+P)

GWM+P relied on Cogmed Working Memory Training (Pearson, 2019), a computerized visuospatial and verbal WM training program (users can modify training session length, number of training days, and weeks). Its robot- and space-themed interface is videogame-like. Its eight games require children to replicate a sequence of events from memory. Three activities involve numbers; one letters. See Supplemental File Table 1 (SFT 1) for description of activities involving numbers and letters. Embedded demonstrations teach the games. Difficulty level, which increases with more targets, more complex tasks, and longer training sequences, is set automatically based on the child's prior success level. Cogmed-certified tutors individually supervised each session as they tracked results and provided support. Each GWM+P session comprised 25 min of Cogmed and 5 min of math practice. Note that in six Progress Indicator sessions distributed across 45 sessions (a standard part of Cogmed), children complete a 1-min multiple-choice single-digit addition test (8 additional seconds of math practice per session).

We adapted Cogmed in two ways. First, we provided timed math practice in arithmetic and WPS (as described), as the final 5 min of each Cogmed session to encourage children to apply increasing WM to math problems. (The same practice sheets were used in all three training conditions.) After children completed the practice problems, tutors provided correct answers, but without instruction on strategies to find correct answers (as was provided in the two contrasting training conditions, where math instruction was an intervention component). Second, we included a self-regulation system (as described) to foster engagement with the WM games and math practice. This incorporated a Robot theme. See Seethaler and Fuchs (2020) for a manual on how Cogmed was implemented in this study.

WPS Intervention

WPS intervention was identical in both study WPS conditions (WPS; WPS+MWM) except for a 6-min game activity, which we describe in the next section. The WPS intervention, known as *Pirate Math*, addresses the dominant second-grade word-problem types in challenging ways, with missing quantities in all three positions of number sentences, 2-digit numbers, irrelevant numbers, relevant numbers provided in graphs and tables, and questions posed in nonstandard fashion. In these ways, it involves flexible problem solving. *Pirate Math* incorporates a pirate theme (e.g., pirate-themed sticker charts; gold coin manipulatives). See Fuchs, Seethaler, et al. (2019) for the manual.

Because research syntheses indicate the importance of structured instruction for improving at-risk children’s learning (Baker et al., 2002; Gersten et al., 2009), WP intervention takes this approach. In the present study, WPS instruction (a) ensures students have the foundational knowledge and skills to succeed with new content; (b) provides explanations in simple, direct language; (c) models efficient solution strategies instead of expecting students to discover strategies on their own; (d) gradually fades support for correct execution of taught strategies; and (e) provides cumulative practice with interleaved problem sets so students use knowledge and strategies to generate many correct responses, distinguish among problem types, and retain previously taught content. (Problem sets were the same in all three training conditions.)

Each WPS intervention session comprises four segments: strategic, speeded practice on arithmetic problems (4 min); the WPS lesson, in which tutors introduce and review WPS concepts and strategies (15 min); games (6 min); and practice (5 min). In this section, we describe speeded practice, the WPS lesson, and practice. In the next section, we describe the games, which is where the two WPS conditions differed.

With strategic, speeded practice ("Meet or Beat Your Score"), children have 60s to answer arithmetic flash cards. Children are taught to "know the answer right off the bat" (retrieve from memory) if confident; otherwise, use the taught counting strategies. Children answer each presented problem correctly because, as soon as an error occurs, they use the taught counting strategy to derive the correct response. To discourage guessing or careless use of counting strategies, seconds elapse as children execute the strategy as many times as needed to produce the correct answer. In this way, careful but quick responding increases correct responses. Children have a second chance to meet or beat the first score. The day's higher score is graphed.

After speeded practice, tutors conduct the *WPS lesson*. For an overview of lesson content, see SFT 2 for WPS intervention content, which is organized in five units. Note that the WPS lesson segment of each WPS session lasted 15 minutes (only a subset of the ideas and activities shown in SFT 2 was addressed in any given session). Unit 1 (lessons 1 - 9) addresses adding and subtracting concepts, addition and subtraction counting strategies, and solving for a missing number, represented by a blank (e.g., 5 – 2 = __; 5 - __ = 3; __ - 2 = 3). Unit 2 (lessons 10 - 18) focuses on combine (referred to as *total*) problems: combining two or three quantities to make a total (e.g., There are 5 girls on the playground and 3 girls in the yard. How many girls are there?). It also includes instruction on 3-part (3-addend) total problems. Unit 3 (lessons 19 - 27) focuses on compare (referred to as *difference*) problems: comparing larger and smaller quantities to find the difference (e.g., At the picnic, the kids ate 5 hot dogs. They ate 3 hamburgers. How many more hot dogs did they eat than hamburgers?). Unit 4 (lessons 28 - 36) focuses on change problems: increasing or decreasing a start quantity to produce an end quantity (e.g., Gamar baked 6 cookies. Then, he gave 3 of them to his friend. How many cookies does Gamar still have?). Unit 5 (37 - 45) provides review and practice.

Units 2 - 4 begin by teaching the mathematical structure of that unit's focal WP type. This involves role playing the problem type's central mathematical event using an intact number story (no missing quantity), concrete objects, and the child's and tutor's names. Tutors next use the intact story to connect the central mathematical event to (a) a visual schematic (into which story quantities are written) and (b) a hand gesture, which is used across lessons to quickly remind children of that problem type's central event. Then tutors connect the WP type's central event to a problem-model number sentence: for total: P1 + P2 = T (Part 1 plus Part 2 equals Total); for difference: B – s = D (bigger quantity minus smaller quantity equals difference); for change: St +/- C = E (for change increase, start number plus change number equals end number; for change decrease, start number minus change number equals end number).

Tutors finally introduce a word problem (with a missing quantity), using the same cover story with which the problem type is introduced. The problem is role played with concrete objects and the child's and tutor's names; the problem type's schematic and hand gesture are applied; and the problem model number sentence is introduced with a blank (__) representing the missing quantity. In each word-problem unit, the idea of irrelevant information is taught.

In Units 2 - 4 for both WPS intervention conditions, tutors teach step-by-step strategies. This includes strategies for understanding problems as belonging to word-problem types (schemas or problem models) and for building the problem model. As children process and solve a problem, they use these strategies to name the problem type, to represent that problem model with the model number sentence, to enter relevant quantities from the problem statement into the problem model while crossing out "extra" (irrelevant) numbers, and to solve for the missing quantity. To foster quick identification of problem types, children play a sorting game where students decide the problem type to which problems belong.

Children RUN through the problem: Read it, Underline what the problem is mostly about (the problem's object code, which becomes the label), and Name the problem type. They write T, D, or C next to the problem to help them remember the problem type, and they write the problem type's model number sentence. Then, they re-read the problem as they enter known quantities and a blank to stand for the unknown quantity into the slots of the problem type's number sentence. For example, given the total problem, There are 5 girls on the playground and 3 girls in the yard. There are also 4 boys in the classroom. How many girls are there?, tutors read aloud the problem as the child follows along. The child underlines *girls*; identifies the problem as total and writes T and the problem type's model number sentence, P1 + P2 = T; re-reads the problem while replacing 5 for P1, 3 for P2, and a blank for T; and crosses out 4. To solve 5 + 3 = ___, the child retrieves the answer or uses counting strategies. See Supplemental File Figure (SFF) 1 for sample second-grade WP work using these strategies.

In each lesson's *5-min practice segment*, children complete the same 2-sided problem set that includes addition and subtraction problems and a word problem (as described and also used in GWM+P). Tutors correct the work (as in GWM+P) and provide elaborated feedback on up to three problems to repair solution strategies (elaborated feedback differs from GWM+P).

WPS Intervention with and without Working Memory Training

In both WPS intervention conditions (with and without WM training), tutors conducted a 5-min game activity between the WPS lesson and practice segment. Games differed between WPS conditions. In WPS intervention, games were designed to reinforce math ideas, skills, and strategies taught during WPS lessons. In WPS+MWM, games were designed to train WM using math stimuli; in both conditions, 55.3% of the games involved word problems. For example, in the "Biggest Number Game," students in the WPS intervention condition identify the operation and the location of the biggest number in an addition or subtraction sentence and then explain why it makes sense that the biggest number is first or last. In WPS+MWM, students do the same but, at the end of the game, they also recall the string of biggest numbers across the items in that 5-min game session. See SFT 3 for an outline of games in the two conditions and the weeks in which they were played.

Training and Support for Tutors

Full or part-time tutors employed by the research project served as tutors. Almost all were university master's students. Less than 10% were certified to teach. Each tutor worked with six students distributed across the three treatment conditions. Tutors were introduced to the intervention conditions in initial workshops. The WPS intervention workshop included an overview of the program; distinctions between the two WPS intervention conditions and methods to ensure students received the condition consistent with random assignment; modeling of key elements; practice implementing elements; explanations and demonstrations of methods for implementing the self-regulation system and providing corrective feedback. For Cogmed, the program publisher provided training online.

After the workshop but prior to the first lesson with a child, tutors completed a reliability quiz covering major components of the intervention conditions with at least 90% accuracy and then demonstrated at least 90% accuracy implementing lesson components with project coordinators. To promote fidelity, tutors studied lesson guides to support their understanding of the lessons. They were not permitted to read the guides while working with children.

In all three conditions, tutors attended weekly meetings in which they provided updates on their students, discussed learning and behavior challenges, and problem-solved with each other, project coordinators, and the first author. Key information on upcoming topics was reviewed and upcoming materials were distributed. Also, every intervention session in all three conditions was audio-recorded. Each week, staff listened to a randomly selected lesson and completed a live observation for each tutor. The purpose was to identify difficulties with or deviations from the protocol, provide quick corrective tutor feedback, and solve problems.

Fidelity of Implementation

Audio-recordings were also used to quantify fidelity of implementation. We sampled 15% of tapes to ensure comparable representation of intervention conditions, tutors, and lesson types. Research assistants and project coordinators independently listened to tapes while completing a checklist to identify essential points addressed in lessons. Coding agreement exceeded 95%. The mean percentage of points addressed in lessons was 98.22 (*SD* = 2.52) in GWM+P; 96.26 (*SD* = 2.67) in WPS+MWM; and 96.49 (*SD* = 2.74) in WPS.

School-Provided WPS Instruction and Mathematics Instructional Time

On a questionnaire completed in the spring, classroom teachers described WPS classroom and intervention instruction. On a 5-point scale, teachers reported how often the school program relied on various sources and methods to teach WPS (1 = never; 2 = rarely; 3 = every once in a while; 4 = sometimes; 5 = a lot). State standards guided instruction, which specified adding and subtracting within 100 with unknowns in all positions involving total, compare, and change situations and solving multi-step problems. Teachers did not, however, teach WPS in terms of problem types (1.42, *SD* = 0.41) and rarely relied on the textbook (1.58, *SD* = 0.84). Reliance on graphic representations was occasional (2.70, *SD* = 0.48), and reliance on a meta-cognitive attack strategy occasionally to sometimes (3.67, *SD* = 1.53). Major instructional foci were labeling answers (4.57, *SD* = 0.79), keywords (4.67, *SD* = 1.31), drawing pictures (4.87, *SD* = 0.53), using objects (4.88, *SD* = 0.46), and using number sentences (4.87, *SD* = 0.56).

Thus, the school program addressed the same word-problem types, with similar focus on labeling and using objects and number sentences to represent word problems. In contrast to study conditions, the school program provided major emphasis on drawing pictures and keywords, which are not deemed productive instructional methods (Powell & Fuchs, 2018). Also, the school program did not rely on a structured approach, and teachers designed their own WP instructional methods with minimal guidance from textbooks.

Across classroom core math instructional time, math center time, and school or study provided math intervention time, math instructional time for control group and intervention students was similar. The mathematics block averaged 347.83 (*SD* = 136.44) min per week, with 79.12 (*SD* = 52.43) min per week allocated to WPS. The supplemental intervention block was 45 - 60 min five times per week. When participating in the study's interventions, 78% of students missed core math instruction or math centers; 17% missed the school's intervention block; 5% missed a different activity. In the control group, 58% of students received school-provided math intervention (mean 154.38 min [*SD* = 62.78] per week); 20% of students in the study's intervention conditions received school-provided math intervention (mean 62.81 min [*SD* = 23.47]) per week.

Procedure

In August, we screened students for study entry. In September - October, we conducted pretesting individually and in small groups. Intervention began in late October and continued through March. One to three weeks after intervention ended, we conducted posttesting individually and in small groups. Four to six weeks after posttest, we administered delayed posttesting in small groups. Testers were blind to study condition when administering and scoring tests. All testing sessions were audio-recorded; 15% of testing tapes were randomly selected, stratifying by tester, which were checked for accuracy by an independent scorer. Agreement exceeded 99%.

Transparency and Openness

This report provides the basis for participant exclusions and the approach used to calculate sample size; identifies the excluded Word Problem WM primary measure; and describes data manipulations and analyses. This report's data are available from the first or third author; the data analysis code is available from the third author; and research materials are available from the first or fifth author. For an unpublished version of this report, which provides results on each indicator of the latent WM factor (including the excluded the Word-Problem WM task), contact the first or third author. This study's design and analysis were not preregistered.

Data Analysis

The study's data structure involved three levels of nesting: 240 students nested in 107 classrooms nested in 16 schools. To account for lack of independence at the classroom level, we used multilevel structural equation models (e.g., Silva et al., 2019) and multilevel models. To account for lack of independence at the school level, we employed a design-based adjustment to standard error computations (Sterba, 2009), which uses a sandwich estimator (Taylor linearization) to account for nonindependence at highest-level units (see TYPE=COMPLEX and CLUSTER IS commands in M*plus* 8.4; Muthén & Muthén, 1998 –2020). We chose a design-based correction for school-level nesting while fitting two-level models, instead of incorporating a third level into models because, for some models, the number of estimated parameters could exceed the number of schools. Two-level models (level 1 = student; level 2 = classroom) were fit in Mplus using full-information robust maximum likelihood estimation (MLR). Note that we did not include study cohort in models because effects were trivial. Also note that pretest WM as a potential moderator of effects is considered in a forthcoming report.

Our first set of analyses involved assessing the conditional main effects of intervention conditions, controlling/adjusting for pretest performance (all fit in Mplus 8.4), using the models in Figures 1a and 1b. In the Figure 1a model, the outcome is *latent WM post*, where posttest WM outcome was treated as a latent variable at the student-level with three manifest indicators: posttest visuospatial WM (VS_WM), posttest verbal-numerals WM (N_WM), and posttest verbal-sentences WM (S_WM). Because hypotheses of interest were at the student-level, classroom-level components of VS_WM, N_WM, and S_WM (i.e., their random intercepts) were simply allowed to covary, in the Figure 1a model entailing a saturated model at the classroom level (with no constraints imposed). To identify the student-level latent WM post factor, one loading was fixed to 1.0, and the factor intercept was fixed to 0. With three indicators and no equality constraints on residual variances, this student-level latent WM posttest factor was just-identified (i.e., model fit indices for the measurement [factor] model necessarily indicate perfect fit). In the Figure 1b model, the outcome *Y post* is a manifest (observed) math outcome variable (posttest arithmetic, delayed-posttest arithmetic, posttest WPS, or delayed-posttest WPS), with the Figure 1b model fit four times.

In Models 1a and 1b, the manifest math outcomes were regressed on their respective (class-mean-centered) pretest covariates. In Models 1a and 1b, D_1, D_2, and D_3 are, respectively, dummy codes for treatment effects for GWM+P versus CON, WPS+MWM versus CON, and WPS versus CON, and their slopes are interpretable as mean differences in latent WM posttest after partialing out pretest on the indicators. To assess mean differences between GWM+P versus WPS+MWM, we used the MODEL CONSTRAINT command to test the difference in the slope of D_1 and the slope of D_2; to compare GWM+P versus WPS, the difference in the slope of D_1 and the slope of D_3 was tested; to compare WPS+MWM versus WPS, the difference in the slope of D_2 and slope of D_3.

Our second set of analyses tested whether latent posttest WM mediated the effect of significant intervention contrasts on the delayed posttest arithmetic or the delayed posttest WPS outcome (controlling for pretest scores on the outcome and controlling for pretest scores on the posttest WM mediator indicators). These intervention contrasts were the statistically significant "total" effects identified in the main effects analyses from Figure 1b. Mediation was investigated using the general multilevel mediation model depicted in Figure 2. In the conventional X→M→Y mediation pathway, *X-variable* corresponds in Figure 2 to a given treatment contrast;

M-variable mediator is latent posttest WM; and *Y-variable* (*delayed-posttest-Y*) is delayed posttest arithmetic or delayed posttest WPS, with Y-*pre* respectively pretest arithmetic or WPS.

In the Figure 2 model, for a given mediation pathway, the *a-path* is the effect of a given treatment contrast on the mediator (latent posttest WM); the *b-path* is the effect of latent posttest WM on the delayed-posttest (arithmetic or WPS) outcome. The *direct effect* is the effect of that same treatment contrast on the delayed-posttest (arithmetic or WPS) outcome. The *indirect effect* (mediation effect) is computed as (*a-path* × *b-path*) and tested for significance when its corresponding total effect *c-path* (treatment contrast on Y-delayed-posttest) was significant in the Figure 1b main effects model. Indirect effects were tested for significance by investigating whether 0 lies within the indirect effect's Monte Carlo 95% confidence interval (Preacher & Selig, 2012) using the utility at www.quantpsy.org. This Monte Carlo confidence interval provides 2.5th and 97.5th percentiles of the sampling distribution of the indirect effect, simulated using parameter estimates from the Figure 2 model.

Results

See Table 1 for pre-, post-, and delayed-posttest means and *SD*s by for students with MD by study conditions (and for classmates). See Table 2 for performance gaps for students with MD against classmates (Cohen's *d* using the classmate *SD*) on outcome measures. See Table 4 for correlations among WM, WPS, and arithmetic measures for the control group of students with MD and for classmates (we did not include students who received intervention, which was designed to disturb relations among WM and math). The .50 mean correlation between WM and WPS was stronger than the .40 mean correlation between WM and arithmetic, $z = 1.90$, 1-tailed $p = .029$ for dependent correlations. This echoes the pattern in the Peng et al. (2016) meta-analysis.

Intraclass correlations at the classroom- and school-level are provided in SFT 4. At the classroom level, these ranged from .023 to .048 for WM indicator measures, .006 to .070 for arithmetic, and .006 to .024 for WPS; at the school level, from .005 to .058 for WM measures, .127 to .156 for arithmetic, and .001 to .003 for WPS.

Main Effect Intervention Contrasts on WM, WPS, and Arithmetic Outcomes

Model fit is calculable for the Figure 1a multilevel structural equation model (MSEM, an overidentified model) but not for the Figure 1b standard multilevel model (MLM, a just-identified model). For the Figure 1a model, model fit is relevant to evaluate at the within-level (student-level), where all pathways of interest were tested (the between-level/classroom-level submodel was saturated). The student-level_ specific fit index, SRMR, was .073 (conventionally, values ≤ .08 indicate adequate fit).

Results of main effects multilevel models along with intervention contrast ESs (adjusted analog to Cohen's *d*) are shown in Table 5. See Figure 3 for a summary of results. On posttest latent WM, general WM training with contiguous math practice (GWM+P) outperformed the other three conditions, and WPS intervention without WM training (WPS intervention) outperformed the control group (CON). On posttest arithmetic, all three intervention conditions outperformed CON, but scored comparably to each other. On delayed posttest arithmetic, significant effects favoring all three intervention conditions over CON persisted. On posttest and delayed-posttest WPS, all three intervention conditions outperformed CON; also, WPS+MWM and WPS intervention outperformed GWM+P.

Posttest WM as a Mediator of Significant Delayed-Posttest Intervention Contrasts

For each significant total effect on a delayed-posttest arithmetic outcome or delayed-posttest WPS outcome, we assessed mediation of the significant total effect (the *c*-path in mediation). In all, there were eight significant intervention contrasts on the delayed posttest arithmetic or delayed-posttest WPS outcomes, with eight mediation pathways tested. For delayed-posttest arithmetic outcome, we assessed mediation of three significant contrasts: GWM+P versus CON, WPS+MWM versus CON, and WPS intervention versus CON; for delayed-posttest WPS, GWM+P versus CON, WPS+MWM versus CON, WPS versus CON, GWM+P versus WPS+MWM, and GWM+P versus WPS intervention.

For each of these mediation pathways, results for the indirect (mediated) effect, the 95% confidence interval, direct effect, *a*-path, and *b*-path are shown in Table 6. Note that for each mediation pathway, the total effect (*c*-path) can also be computed as the direct effect plus the indirect effect, and the *b*-path estimate is identical for analyses with the same outcome and same treatment contrast (we repeatedly report the *b*-path estimate for clarity).

The three significant indirect effects all pertained to total effects on delayed-posttest WPS, revealing the following. The GWM+P versus CON effect on delayed-posttest WPS was significantly mediated by the latent WM posttest factor (1.019* 95%CI = [0.048, 1.792]) as were the WPS intervention versus CON effect on delayed-posttest WPS (0.532* 95% CI = [0.018, 1.238]) and the GWM+P versus WPS+MWM effect on delayed-posttest WPS (0.747* 95%CI = [.037, 1.318]).

Discussion

The purpose of this randomized controlled trial was to contrast the effects of three intervention conditions for improving WM and WPS outcomes among students with MD. Two conditions involved WM training, designed in different ways to support transfer of strengthened WM to WPS. One of the two WM training conditions also included WPS intervention. The third condition was WPS intervention without WM training. Results provide the basis for five conclusions. First, general WM training with contiguous math practice (GMW+P) improves WM and WPS outcomes. Second, WM training is not a substitute for WPS intervention when the goal is to strengthen WPS outcomes. Third, WPS intervention without WM training improves WM, but WPS intervention is not a substitute for general WM training when the goal is to strengthen WM. Fourth, for WM effects to accrue, WM training requires sufficient intensity. Fifth, results provide support for WM's causal role in WPS, but not in arithmetic. We next consider findings pertaining to each conclusion. Then, we draw implications for practice and research, including the need to study instructional supports for building cognitive-academic bidirectionality.

General WM Training with Contiguous Math Practice Improves WPS Performance

GWM+P involved a form of general WM training with previously documented effects on WM (Gray et al., 2012; Shinaver et al., 2014), but with questionable evidence of transfer to math (Bergman Nutley & Söderqvist, 2017; Roche & Johnson, 2014). GWN+P's innovation was aimed at supporting transfer from strengthened WM to WPS by providing 5 min of math practice immediately after each 25-min dose of WM training.

Findings extend prior work by demonstrating effects of this approach on WM as well as on WPS outcomes. On complex WM span, GWM+P outperformed the control group as well as the training condition without WM training (i.e., WPS intervention). ESs (see Table 5) on WM were moderate to large: for GWM+P versus CON, 1.22; for GWM+P versus WPS intervention, 0.60. In line with the study's hypothesis, GWM+P also conferred advantage on WPS, a transfer domain not addressed in prior work. The ES for GWM+P versus CON on WPS of 0.32 is notable, in part because it remained significant 4 - 6 weeks later, at delayed posttest (ES = 0.28).

Performance gaps for students in the GWM+P condition also decreased substantially: on visuospatial WM, from nearly 1 standard deviation (*SD*; 0.94) below classroom peers at pretest to 0.50 at posttest and on verbal-numerals WM, from 1.06 to 0.56. By contrast, the control group's WM gaps increased or remained similarly large (1.05 to 1.47 *SD*s below classmates on visuospatial WM; 1.11 to 0.89 on verbal-numerals WM). Normalization on WPS, while more modest (1.11 to 0.84 *SD*s), was notable because CON's WPS gap widened over the same time period (1.07 to 1.16 *SD*s).

These findings suggest that simply providing math practice immediately after each WM training session is a potentially productive approach to support transfer from WM training to WPS. Because feedback on GWM+P's math practice was limited to the provision of correct answers (without structured WPS instruction), it is unlikely that students with MD acquired new WPS skill as a function of this practice. We instead surmise that GWM+P increased WM, which in turn boosted children's capacity to consolidate pre-existing WPS skill during the training sessions' contiguously presented practice or to learn new material from classroom WPS instruction by permitting students with MD to store and process information more efficiently.

This interpretation is supported by this study's mediation analyses, which indicate that improved WPS accrues via strengthened WM capacity. In fact, the WM latent factor fully mediated GWM+P's effects on delayed-posttest WPS (i.e., the direct path from intervention to WPS was no longer significant with the indirect effect in the model). This is consistent with the study's hypothesis. Our results extend the literature by demonstrating that general WM training with contiguous math practice improves WM capacity and transfers to WPS for students with MD.

Such General WM Training Is Not a Substitute for WPS Intervention

This prompts the question: Is such WM training a substitute for WPS intervention? Study findings contraindicate this idea, because both WPS conditions (with and without WM training) convincingly and significantly outperformed GWM+P on the WPS outcome. The posttest ES for GWM+P versus WPS was 0.97; for GWM+P versus WPS+MWM, 0.71. Moreover, advantages for both WPS intervention conditions on WPS at delayed posttest remained strong (ESs = 1.49 and 0.74).

This finding is not surprising given that WPS intervention provided intensive, structured instruction to teach productive strategies for building word-problem models. In fact, although the pretest WPS performance of students with MD placed them 1 *SD* below classmates, the WPS performance gap in the WPS intervention (without WM training) condition was more than completely closed at posttest, with students with MD performing 0.94 *SD*s above classmates.

Present study findings therefore indicate that although general WM training with contiguous math practice transfers to and confers advantage on WPS, compared to the control group, direct WPS intervention is a more efficient approach for improving WPS than is general WM training with math practice. This finding is important in the context of school settings, where time for supplemental intervention is limited.

WPS Intervention Improves WM

Including the WPS intervention condition (without WM training) permitted us to assess whether WPS intervention improves WM. Within and across studies, findings to date on whether math intervention improves WM are mixed. A few prior studies have tested this idea when training other math domains. Nelwan and Kroesbergen (2016) and Nelwan et al. (2018) found no effects on WM during arithmetic training. This was also the case for a counting-on numerical magnitude comparison game (Ramani et al., 2019), although a subsequent study provided more promising results (Ramani et al., 2017). Moore et al. (2012) obtained mixed findings when assessing the effects of a broad-based math intervention, including calculations and WPS instruction, in children with leukemia undergoing central nervous system treatment.

Our findings are in line with prior studies supporting the idea that math training can strengthen WM. In the present study, WPS intervention significantly outperformed CON on WM, with an ES of 0.62. This is a relatively large ES in the context of prior studies, perhaps because this study's WPS instruction more transparently maps onto WM processes than does arithmetic or magnitude instruction. Even so, WPS intervention's effect on WM (ES = 0.62) was considerably lower than general WM training's effect on WM (1.22), and the effect between these conditions was significant (ES = 0.60). This indicates that WPS intervention is not a substitute for general WM training when the goal is to build WM, as might be the case when attempting to build capacity across performance domains.

Results nevertheless suggest that academic skills training, when focused on a math domain with a relatively strong association with WM (Peng et al., 2016) and when infused with WM-rich strategic activities, enhances WM capacity, even as it confers strong advantage on WPS. Findings thus suggest that the present study's WPS intervention, which teaches strategies to support updating processes thought to be engaged during WPS (see the account based on Kintsch & Greeno, 1985), represents a form of WM training.

Additionally, mediation analyses support this study's hypothesis that WPS intervention strengthens complex WM span, which contributes to WPS intervention's effects on the WPS outcome: On delayed posttest WPS, strengthened WM partially mediated the effect between WPS intervention and CON: WPS intervention improved WM, and each unit of improved WM was associated with stronger WPS outcome. Combined with analyses indicating that WM fully mediated GWM+P's effect on WPS, this suggests that cognitive-academic growth may be bidirectional, as framed by Peng and Kievit (2020). We revisit this point later in this discussion.

WM Training Needs to Occur with Sufficient Intensity

This study's remaining training condition combined the same WPS intervention with domain-specific WM training (WPS+MWM). This was designed to confer WPS advantage over the other training arms. We expected stronger WPS for WPS+MWM over general WM training with math practice for two reasons. First, consistent with transfer-appropriate processing theory (Franks et al., 2000; Roediger et al., 1989), we expected it to strengthen congruency between processes used during learning and those required in the targeted outcome. Second, it incorporated WPS intervention to build WPS strategies, onto which students with MD might apply strengthened WM. At the same time, we also hypothesized stronger WPS outcome for WPS+MWM over WPS intervention without WM training, given expectations that domain-specific WM training strengthens WM, which improves WPS by permitting students with MD to store and process information more efficiently during WPS intervention's WPS instruction.

Unfortunately, the WPS+MWM condition failed to strengthen WM. The WPS+MWM versus CON contrast on the latent WM factor was not significant (ES = 0.30), and analyses revealed that the significant effect between WPS+MWM and CON on the WPS outcome was not mediated by strengthened WM. By contrast, the significant effect between WPS+MWM versus GWM+P, which favored the general WM training with math practice condition, was mediated by strengthened WM in the general WM training condition. This pattern of findings casts doubt on whether the WPS+MWM intervention improved WM and whether significant effects involving the WPS+MWM intervention on the WPS outcomes can be attributed to the domain-specific WM training component. Instead, findings suggest that the WPS+MWM intervention's effects on WPS outcomes reside entirely with the WPS component of this combined intervention.

Prior studies also raise questions about the tenability of domain-specific WM training involving numerical stimuli. Although Rode et al. (2014) found that such WM training produced improvement on the WM training task as well as untrained measures of verbal and visuospatial WM, transfer to math was mixed: A small significant effect occurred on one measure (ES = 0.26) but not the other (ES = 0.05). Kroesbergen et al. (2014) contrasted three conditions: a control group and WM training with and without numerical stimuli. On phonological WM, neither WM training group outperformed control; on visuospatial WM, both training conditions outperformed control, but without differences between training conditions. On counting skills, domain-specific but not general WM training outperformed control, but the training conditions performed comparably. Still, Ramani et al. (2017) identified positive WM effects. Given differences among these and other prior studies, the conditions that support positive effects for domain-specific WM training remain unclear.

In terms of the present study's domain-specific WM training condition's absence of convincing WM effects, it is important to note that, to control for instructional time across this randomized controlled trial's training arms, the WPS+MWM condition's WM training time was necessarily dramatically less than in the general WM condition (GWM+P): In WPS+MWM, 6 min per session in 38 sessions (for a total of 228 minutes; games are played in 38 of the 45 session in WPS+MWM and in WPS); in GWM+P, 25 minutes per session in 45 sessions (for a total of 1,125 minutes). Accordingly, on the latent WM factor, GWM+P but not WPS+MWM significantly outperformed CON (respective ESs = 1.20 vs. 0.30).

In a similar vein, finding that WPS intervention but not WPS+MWM improved WM outcomes suggests that reduced time on the academic component of the WPS+MWM intervention is also problematic. Although a reduction of 6 min WPS instructional time may seem inconsequential (228 min over 38 sessions), WPS without WM training games addressed fluency development on key WPS ideas and strategies, with many repetitions provided within each 6-min game. Such repetition was not possible in the WPS+MWM games due to the time spent recalling. Further, the WPS intervention without WM training condition's fluency development may reduce cognitive load during WPS in a way that permits engagement of updating processes during WPS, thereby simultaneously strengthening WM and WPS. This hypothesized mechanism, which warrants future study, suggests that the reduction in both WPS instructional time and WM training time may have reduced the combined WPS+MWM condition's impact on WM and WPS outcomes.

A Causal Role for WM in WPS, Not in Arithmetic

A major contribution of the present study within the WM training literature is its novel focus on WPS. Given the pattern of correlations in Peng et al.'s (2016) meta-analysis (a pattern corroborated in the present study with stronger correlations between WM and WPS and between WM and arithmetic), we thought it possible that WPS, a form of complex mathematical reasoning, may provide greater opportunity for transfer from WM training than is the case with the less complex math transfer targets addressed in prior work. To gain insight into this possibility, we also included an outcome measure on arithmetic, the dominant outcome in the

WM training literature. We hypothesized that strengthened WM mediates significant condition effects on WPS but not on arithmetic.

All three training conditions outperformed the control group on arithmetic, with no significant differences among the training arms. At posttest, the ES (see Table 5) for the significant contrast between GWM+P and CON was 0.64; between WPS intervention and CON, 0.71; and between WPS+MWM and CON, 0.81. At delayed posttest, ESs diminished (respectively, to 0.57, 0.45, and 0.32) but the contrasts remained significant. Most interestingly, and consistent with this study's hypothesis, strengthened WM did not mediate any of these effects. This was largely due to the nonsignificant mediation *b*-path from strengthened WM to arithmetic.

By contrast, strengthened WM mediated three of the five tested (significant) treatment contrasts: for GWM+P versus CON, for WPS intervention versus CON, and for GWM+P versus WPS+MWM. One of the two nonsignificant indirect (mediation) effects involved the WPS+MWM versus CON contrast which, as discussed, reflects WPS+MWM's failure to strengthen WM. This is indicated in the mediation *a*-path. The other nonsignificant indirect (mediation) effect was for the GWM+P versus WPS contrast. Here, too, the absence of significant mediation effect reflects the *a*-path: Although the GWM+P intervention outperformed the WPS intervention on WM, the difference between these conditions was weaker than for other contrasts.

The specificity of these mediation results, in which mediation was supported for the WPS outcome but not the arithmetic outcome, increases confidence when concluding that the arithmetic effects reported in the present study are due to the training conditions' math components, not to strengthened WM. Further, this randomized controlled trial's treatment effects and mediation pathways together provide evidence for a causal role for WM in WPS, but not in arithmetic. This may reflect WPS's demands on a greater array of cognitive resources than is involved in arithmetic, including WM (e.g., Fuchs et al., 2018).

Some readers might nevertheless wonder whether arithmetic training or arithmetic training that incorporates word problems in meaningful contexts, but without specific WPS instruction, might be sufficient to improve WPS. Prior work suggests this is not the case. For example, a recent randomized controlled trial (Fuchs et al., 2021) showed that WPS outcomes were dramatically superior for students who received WPS intervention compared to those who received arithmetic and number knowledge training without WPS intervention (Fuchs et al., 2021). Further, in an earlier experiment, students who received arithmetic training that incorporated word problems in meaningful contexts lost ground relative to classmates even as their arithmetic achievement gap narrowed (Fuchs et al., 2013). These studies suggest the importance of dedicated WPS intervention.

Limitations

Before discussing implications for research and practice, we draw readers' attention to three important study limitations. First, we note that a causal role for WM in WPS but not arithmetic may be specific to second grade. Relations may differ in other age groups or as a function of how arithmetic is assessed and how WPS is operationalized and measured. Answering questions about changing relations across development and task demands requires additional studies. Second, although causal inference from the mediation pathways is facilitated in the present study because conditions were randomly assigned and due to the temporal precedence of measurements (i.e., measurement of covariates preceded random assignment to treatments, which preceded treatment administration, which preceded measurement of the mediator, which preceded measurement of the delayed-posttest outcomes), mediation analysis is essentially correlational.

A third limitation is that the present study's focus was confined to the role of complex WM span within WPS, even though other domain-general resources are involved in WPS. For example, in an individual differences study, Majumder (2004) demonstrated that attentional inhibition also plays a role in the solution of some math word problems (while controlling for the effects of WM) and that reading comprehension provides added value in predicting WPS. Further, in a recent randomized controlled trial, oral language comprehension was identified as playing a causal role in WPS (Fuchs et al., 2021).

To isolate the contribution of complex WM span, future studies focused on the role of complex WM span would benefit from controls for these and other variables thought to be involved in WPS. Studies contrasting the efficacy of WM training with training on inhibition or reading comprehension or language would also make a valuable contribution, and contrasting WM training to a condition focused on a more comprehensive set of executive functions may be informative. Additionally, it would be interesting to consider whether training on complex WM span affects related domain-general processes, such as inhibition, or other academic skills involved in WPS, such as reading comprehension.

Implications for Practice and Other Research: A Focus on Building Cognitive-Academic Bidirectionality

With these limitations in mind, we draw the following implications from the present study's findings for practice and other future research. First, general WM training with a simple, efficient priming procedure supports transfer to math, but it should not be viewed as a substitute for intervention designed specifically to improve performance in complex academic domains. Second, academic interventions whose instructional components naturally incorporate updating and other strategies that call upon WM may improve WM in the absence of specific WM training. Together, these conclusions offer promise for the idea of building cognitive-academic bidirectionality, perhaps by marrying cross-modality cognitive training with academic skill building, while ensuring sufficient intensity on both components.

Finally, a successful marriage may benefit from an instructional focus on transfer designed to explicitly connect the cognitive and academic components via a meta-cognitive strand that is embedded within the cognitive and academic components. Such a 3-pronged approach may represent a promising direction in future research for building cognitive-academic bidirectionality. This may be important for addressing comorbid forms of learning difficulty, given that WM resources are used across domains of academic learning.

Using WM and WPS as an example, one might test the effects of intervention that combines general, cross-modality WM training with a validated WPS intervention via explicit transfer instruction. As in previous work within the domain of WPS (Fuchs et al., 2003), explicit transfer instruction would teach children what *transfer* means; help them appreciate the commonalities between the cognitive demands involved in WPs and WM games; and, following Fuchs et al., Jones et al. (2020), and Partanen et al. (2015), encourage meta-cognitive engagement, with oral and visual reminders to apply WM capacity during WPS. In fact, the effects of cognitive training on executive skills are stronger with the addition of metacognitive scaffolding (Pozuelos et al., 2018), and such scaffolding has been suggested as important for increasing executive functions and closing achievement gaps, especially in mathematics (Zelazo & Carlson, 2020).

Testing the effects of such innovations would require analyses that assess cognitive-academic bidirectionality. It would also require a multi-arm study that assesses the added value of the novel instructional component (WM training with explicit transfer instruction and meta-cognitive engagement) when the base program (WPS intervention) is intact versus when it is reduced to control for instructional time. Such a design would help researchers disentangle the substantive value of the 3-pronged approach from the effects of extra instructional time. The larger goal would be to deepen insight into methods for fostering cognitive-academic bidirectionality by simultaneously building learning capacity and academic competence in children.

References

Alloway T. P. (2012). *Alloway Working Memory Assessment 2 (AWMA-2).* London: Pearson.

Baker, S., Gersten, R., & Lee, D-S. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *The Elementary School Journal*, 103, 51-73.

Batty, G. D., Kivimäki, M., & Deary, I. J. (2010). Intelligence, education, and mortality. *BMJ: British Medical Journal, 340,* 1-2. doi: http://dx.doi.org/10.1136/bmj.c563

Bergman Nutley, S., & Söderqvist, S. (2017). How is working memory training likely to influence academic performance? Current evidence and methodological considerations. *Frontiers in Psychology, 8,* Article 69. https://doi.org/10.3389/fpsyg.2017.00069

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

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

Every Child a Chance Trust (2009). *The long-term costs of numeracy difficulties*. Retrieved. August 14, 2009, http://www.everychildachancetrust.org/counts/index.cfm.

Franks, J. J., Bilbrey, C. W., Lien, K. G., & McNamara, T. P. (2000). Transfer-appropriate processing (TAP) and repetition priming. *Memory & Cognition, 28*, 1140-1151.

Fuchs, L.S., Fuchs, D., Craddock, C., Hollenbeck, K.N., Hamlett, C.L., & Schatschneider, C. (2008). Effects of small-group tutoring with and without validated classroom instruction on at-risk students' math problem solving: Are two tiers of prevention better than one? *Journal of Educational Psychology, 100*, 491-509.

Fuchs, L.S., Fuchs, D., Prentice, K., Burch, M., Hamlett, C.L., Owen, R., Hosp, M., & Janeck, D. (2003). Explicitly teaching for transfer: Effects on third-grade students' mathematical problem solving. *Journal of Educational Psychology, 95*, 293-304.

Fuchs, L. S., Fuchs, D., Seethaler, P. M., & Barnes, M. A. (2020). Addressing the role of working memory in mathematical word-problem solving when designing intervention for struggling learners. *ZDM Mathematics Education, 52*, 87-96. doi: 10.1007/s11858-019-01070-8.

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., Seethaler, P. M., Bryant, J. V., & Schatschneider, C. (2010). Do different types of school mathematics development depend on different constellations of numerical and general cognitive abilities? *Developmental Psychology, 46*, 1731-1746. doi: 10.1037/a0020662

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Schatschneider, C., Hamlett, C. L., DeSelms, J., Seethaler, P. M., Wilson, J., Craddock, C. F., Bryant, J. D., Luther, K., & Changas, P. (2013). Effects of first-grade number knowledge tutoring with contrasting forms of practice. *Journal of Educational Psychology, 105*, 58-77.

Fuchs, L. S., Gilbert, J. K., Fuchs, D., Seethaler, P. M., & Martin, B. (2018). Text comprehension and oral language as predictors of word-problem solving: Insights into word-problem solving as a form of text comprehension. *Scientific Studies of Reading, 22,* 152-166. doi.org/10.1080/10888438.2017.1398259

Fuchs, L.S., Seethaler, P.M., Sterba, S. Y., Craddock, C., Fuchs, D., Compton, D.L., Geary, D.C., & Changas, P. (2021). Closing the word-problem achievement gap in first grade: Schema-based word-problem intervention with embedded language comprehension instruction. *Journal of Educational Psychology, 113*, 86-103. DOI: 10.1037/edu0000467

Fuchs, L.S., Seethaler, P.M., & Fuchs, D. (2019). *First-Grade Pirate Math Manual*. Available from L.S. Fuchs, Vanderbilt University, Nashville, TN 37203.

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research, 79*, 1202–1242. doi: 10.3102/0034654309334431

Gray, S. A., Chaban, R., Martinussen, R., Goldberg, H., Gotlieb, R., Kronitz, R., Hockenberry, M., & Tannock, R. (2012). Effects of a computerized working memory training program on working memory, attention, and academics in adolescents with severe LD and comorbid ADHD: A randomized clinical trial. *Journal of Child Psychology and Psychiatry, 53*, 1277–1284. doi: 10.1111/j.1469-7610.2012.02592.x

Haskell, R. E. (2001). *Transfer of learning: Cognition, instruction, and reasoning*. Academic Press. https://doi.org/10.1016/B978-012330595-4/50003-2

Hoffer, T. B., Venkataraman, L., Hedberg, E. C., & Shagle. S. (2007). *Final report on the National Survey of Algebra Teachers for the National Math Panel* (retrieved 3/26/2019). https://www.researchgate.net/profile/Eric_Hedberg/publication/228513219_Final_report_on_the_national_survey_of_algebra_teachers_for_the_National_Math_Panel/links/55364e710cf20ea35f11ca6d/Final-report-on-the-national-survey-of-algebra-teachers-for-the-National-Math-Panel.pdf.

Houghton-Mifflin Harcourt. (2015 - 2016) Go MATH! https://www.hmhco.com/programs/go-math#overview

Jones, J.S., Milton, F., Mostazir, M., & Adlam, A.R., (2020). The academic outcomes of working memory and metacognitive strategy training in children: A double-blind randomized controlled trial. *Developmental Science*, *23*:e12870. https://doi.org/10.1111/desc.128707

Jordan, N. C., & Hanich, L. (2000). Mathematical thinking in second-grade children with different forms of LD. *Journal of Learning Disabilities, 33*, 567-578. doi:10.1177/002221940003300605

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

Kintsch, W., & Greeno, J. G. (l985). Understanding and solving word arithmetic problems. *Psychological Review, 92*, 109-129.

Koedinger, K. R., & Nathan, M. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *Journal of the Learning Sciences, 113*, 129-164. doi: 10.1207/s15327809jls1302_1

Kroesbergen, E. H., van't Noordende, J. E., & Kolkman, M. E., (2014). Training working memory in kindergarten children: Effects on working memory and early numeracy. *Journal of Child Neuropsychology, 20*, 23-37. DOI: 10.1080/09297049.2012.736483

Li, D., Christ, S. E., & Cowan, N. (2014). Domain-general and domain-specific functional networks in working memory. *Neuroimage, 102*, 646-656. doi: 10.1016/j.neuroimage.2014.08.028

Majumder, S. (2004). Factors in mathematical word problem solving: The role of inhibition. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 64*(8-B), 4077.

Melby-Lervag, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology, 49*, 270-291. doi:10.1037/a0028228.

Moore, I. M., Hockenberry, M. J., Anhalt, C., McCarthy, K., & Krull, K. R. (2012). Mathematics intervention for prevention of neurocognitive deficits in childhood leukemia. *Pediatric Blood & Cancer*, *59*, 278-284.

Muthén, L. K., & Muthén, B. O. (1998-2020). *Mplus User's Guide 8.4*. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B. O. (1998-2013). *Mplus users guide* (7th ed.). Los Angeles: Muthén & Muthén.

Muthén, L.K., & Muthén, B.O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 4*, 599-620.

National Research Council. (2000). *How people learn: Brain, mind, experience, and school*. The National Academies Press.

Nelwan, M., & Kroesbergen, E. H. (2016). Limited near and far transfer effects of Jungle Memory working memory training on learning mathematics in children with attentional and mathematics difficulties. *Frontiers in Psychology*. DOI: 10.3389/fpsyg.2016.01384

Nelwan, M., Vissers, C., & Kroesbergen, E. H. (2018). Coaching positively influences the effects of working memory training on visual working memory as well as mathematical ability. *Neuropsychologia, 113*, 140-149. DOI: 10.1016/j.neuropsychologia.2018.04.002

Partanen, P., Jansson, B., Lisspers, J., & Sundin, Ö. (2015). Metacognitive strategy training adds to the effects of working memory training in children with special educational needs. *International Journal of Psychological Studies, 7* 10.5539/ijps.v7n3p130

Passolunghi, M. C., & Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *Journal of Experimental Child Psychology 88*, 348-367. DOI: 10.1016/j.jecp.2004.04.002

Pearson (2019). *Cogmed Working Memory Training Program*. https://www.cogmed.com/

Peng, P., & Kievit, R. A. (2020). The development of academic achievement and cognitive abilities: A bidirectional perspective, 14(1*), 15-20. Child Development Perspectives*. https://doi.org/10.1111/cdep.12352

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

Pickering, S., & Gathercole, S. (2001). *Working Memory Test Battery for Children.* London: The Psychological Corporation.

Powell, S. R., & Fuchs, L. S. (2018). Effective word-problem instruction: Using schemas to facilitate mathematical reasoning. *TEACHING Exceptional Children, 51*(1), 31-42. doi.org/10.1177/0040059918777250

Powell, S. R., Fuchs, L. S., Fuchs, D., Cirino, P. T., & Fletcher, J. M. (2009). Do word-problem features differentially affect problem difficulty as a function of students' mathematics difficulty with and without reading difficulty? *Journal of Learning Disabilities, 42*, 99-110. DOI: 10.1177/0022219408326211

Pozuelos, J. P., Combita, L. M., Abundis, A., Paz-Alonso, P. M., Conejero, Á., Guerra, S., & Rueda, M. R. (2018). Metacognitive scaffolding boosts cognitive and neural benefits following executive attention training in children. *Developmental Science*, *22*(2), e12756.

Preacher, K. J., & Selig, J. P. (2012) *Advantages of Monte Carlo confidence intervals for indirect effects. Communication Methods and Measures*, *6*(2), 77-98, DOI: 10.1080/19312458.2012.679848

Ramani, G. B., Daubert, E. N., Lin, G. C., Kamarsu, S., Wodzinski, A., & Jaeggi, S. M. (2019). Racing dragons and remember aliens: Benefits of playing number and working memory games on kindergarteners' numerical knowledge. *Developmental Science,* 2019;00;e12908. https://doi.org/10.1111/dewsc.12908

Ramani, G. B., Jaeggi, S. M., Daubert, E. N., & Buschkuehl, M. (2017). Domain-specific and domain-general training to improve kindergarten children's mathematics. *Journal of Numerical Cognition, 3*(2), 488-495. doi: 10.5964/jnc.v3i2.31

Roche, J.D., & Johnson, B.D. (2014). Cogmed Working Memory Training product review. *Journal of Attention Disorders, 18*, 379-384. DOI: 10.1177/1087054714524275

Rode, C., Robson, R. Purviance, A., Geary, D.C., & Mayr, U. (2014). Is working memory training effective? A study in a school setting. *PLOS ONE, 9*(8). https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0104796

Roediger, H. L., Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H.L. Roediger & F.I.M. Craik (Eds.), *Varieties of memory and consciousness: Essays in Honour of Endel Tulving.* Erlbaum.

Sala, G., & Gobet, F. (2017). Working memory training in typically development children: A meta-analysis of the available evidence. *Developmental Psychology, 53*, 671-685. DOI:10.1047/dev0000265

Seethaler, P. M., & Fuchs, L. S. (2020). *Implementing Cogmed with Contiguous Math Practice and a Motivational Component: A Manual.* Available from L. S. Fuchs, Vanderbilt University, Nashville TN 37203.

Shinaver, C. S., Entwistle, P. C., & Söderqvist, S. (2014). Cogmed WM training: Reviewing the reviews. *Applied Neuropsychology: Child, 3*, 163-172. doi: 10.1080/21622965.2013.875314

Silva, B. C., Bosancianu, C. M. & Littvay, L. (2019). Multilevel structural equation modeling. Los Angeles: Sage.

Sterba, S. K. (2009). Alternative model-based and design-based frameworks for inferences from samples to populations: From polarization to integration. *Multivariate Behavioral Research, 44*, 711-740. doi: 10.1080/00273170903333574

Swanson, H. L. (2016). Word problem solving, working memory and serious math difficulties: Do cognitive strategies really make a difference? *Journal of Applied Research in Memory and Cognition, 5*, 368-383. DOI: 10.1016/j.jarmac.2016.04.012

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

Swanson, H. L., Moran, A., Lussier, C., & Fung, W. (2014). The effect of explicit and direct generative strategy training and working memory on word problem-solving accuracy in children at risk for math difficulties. *Journal of Learning Disabilities, 37*, 111-123. DOI: 10.1177/0731948713507264

Swanson, H. L., & Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: Both executive and phonological processes are important. *Journal of Experimental Child Psychology*, *79*, 294-321.

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Psychological Corporation.

What Works Clearinghouse (2009). *Assisting students struggling with mathematics: Response to intervention (RtI) for elementary and middle schools*. U.S. Department of Education, Institute of Education Sciences NCEE 2009-4060. https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/rti_math_pg_042109.pdf

What Works Clearinghouse (2021). *Assisting Students Struggling with Mathematics: Intervention in the Elementary Grades (Educator's Practice Guide).* U.S. Department of

Education, Institute of Education Sciences NCEE 2021-006. https://ies.ed.gov/ncee/wwc/Docs/PracticeGuide/WWC2021006-Math-PG.pdf

Zelazo, P. D., & Carlson, S. M. (2020). The neurodevelopment of executive function skills: Implications for academic achievement gaps. *Psychology & Neuroscience, 13*(3), 273-298. http://dx.doi.org/10.1037/pne0000208

Zhang, H., Chang, L., Chen, X., Ma, L., & Zhou (2018). Working memory updating training improves mathematics performance in middle school students with learning difficulties. *Frontiers in Human Neuroscience*. DOI: 10.3389/fnhum.2018.00154

Zhang, Q., Wang, C., Zhao, Q., Yang, L., Buschkuehl, M., & Jaeggi, S. M. (2019). The malleability of executive function in early childhood: Effects of schooling and targeted training. *Developmental Science*, *22*(2), e12748.

Table 1

Means and Standard Deviations by Study Condition

	MD Study Condition									
	Control (n=57)		GWM+P (n=63)		WPS+MWM (n=60)		WPS (n=60)		Non-MD (n=157)	
Variable	Mean	*(SD)*	Mean	*(SD)*	Mean	*(SD)*	Mean	*(SD)*	Mean	*(SD)*
Working Memory										
Visuospatial: Pre	8.61	(5.89)	9.33	(5.55)	9.28	(4.63)	8.97	(5.42)	15.75	(6.83)
Post	10.96	(5.50)	16.38	(3.92)	12.33	(5.75)	13.48	(6.09)	19.17	(5.60)
Verbal – Numerals: Pre	11.63	(3.95)	11.89	(5.22)	11.95	(4.39)	12.33	(4.34)	16.84	(4.68)
Post	14.07	(4.28)	15.81	(4.08)	14.35	(4.32)	15.10	(3.94)	18.83	(5.36)
Verbal – Sentences: Pre	3.89	(3.01)	4.02	(3.65)	5.08	(4.00)	4.38	(3.03)	9.52	(3.86)
Post	6.28	(4.21)	7.22	(4.20)	7.62	(3.53)	7.72	(3.60)	11.78	(3.28)
Mathematics										
Arithmetic: Pre	11.51	(7.21)	15.11	(10.53)	13.25	(7.99)	12.10	(8.12)	28.16	(13.88)
Post	24.12	(14.51)	36.05	(19.02)	34.98	(14.95)	34.50	(14.04)	46.00	(21.14)
Delayed	4.42	(2.58)	6.13	(2.64)	5.15	(2.51)	5.57	(2.32)	---	---
Word Problems: Pre	2.25	(1.68)	2.10	(1.80)	2.32	(1.72)	2.42	(1.70)	5.99	(3.51)
Post	4.19	(3.07)	6.03	(4.53)	14.98	(7.33)	16.13	(8.13)	10.80	(5.70)
Delayed	2.30	(2.00)	3.27	(2.81)	5.87	(3.89)	6.85	(4.02)	---	---

GWM+P is general working memory (WM) training with contiguous practice; WPS+MWM is word-problem solving (WPS intervention plus math (domain-specific) WM training; WPS is WPS intervention; MD is math difficulties; Visuospatial is *Automated Working Memory Assessment (AWMA; Alloway, 2012)– Odd-One Out*; Verbal-Numerals is *WMTB-C (Pickering & Gathercole, 2001)–Counting Recall; Verbal Sentences* is is *AWMA–Listening Recall; Arithmetic* is a sample of problems drawn from the school program, the district's online scope, and sequence and sample units for state standards, as is the case for *Second-Grade Word Problems*.

Table 2

Performance Gaps by MD Study Condition: Effect Sizes (ESs; Hedges g) for MD Study Conditions Versus Classmates

	MD Study Condition			
	Control (n=57)	GWM+P (n=63)	WPS+MWM (n=60)	WPS (n=60)
Variable	ES	ES	ES	ES
Working Memory				
Visuospatial: Pre	1.05	0.94	0.95	0.99
Post	1.47	0.50	1.22	1.02
Verbal - Numerals: Pre	1.11	1.06	1.04	0.96
Post	0.89	0.56	0.84	0.70
Verbal – Sentences: Pre	1.46	1.42	1.15	1.33
Post	1.68	1.39	1.27	1.24
Mathematics				
Arithmetic: Pre	1.20	0.94	1.07	1.16
Post	1.04	0.47	0.52	0.54
Word Problems: Pre	1.07	1.11	1.05	1.02
Post	1.16	0.84	-0.73	-0.94

GWM+P is general working memory (WM) training with contiguous practice; WPS+MWM is word-problem solving (WPS intervention plus math (domain-specific) WM training; WPS is WPS intervention; MD is math difficulties; Visuospatial is *Automated Working Memory Assessment (AWMA;* Alloway, 2012*)– Odd-One Out*; Verbal-Numerals is *WMTB-C (*Pickering & Gathercole, 2001*)–Counting Recall; Verbal Sentences* is *AWMA–Listening Recall; Arithmetic* is a sample of problems drawn from the school program, the district's online scope, and sequence and sample units for state standards, as is the case for *Second-Grade Word Problems*.

Table 3

Minutes Spent in Training Segments by Condition

	Conditions		
Training Segment	GWM+P	WPS+MWM	WPS
General (visuospatial and verbal) WM games	25	0	0
Speeded strategic arithmetic practice	0	4	4
WPS problem-type and instructional strategies instructional lesson	0	15	15
Arithmetic and WPS games to develop fluency with foundational skills	0	0	6
Arithmetic and WPS games to develop WM capacity	0	6	0
Math practice on arithmetic and WPS			
With feedback accuracy	5[a]	0	0
With feedback accuracy and brief corrective review	0	5	5
Total	30	30	30

GWM+P is general working memory (WM) training with contiguous practice; WPS+MWM is word-problem solving (WPS intervention plus math (domain-specific) WM training; WPS is WPS intervention. [a] GWM+P includes six Progress Indicator sessions, each with a 1-min multiple-choice single-digit arithmetic test. If one counts this as math practice, 5 min is 5.13 min (6 min over the 45-session training = 0.13 min per session or approximately 8 additional seconds per session).

Table 4

Correlations among Pretest and Posttest Working Memory and Math Measures (N=214[a])

	Pretest								
	Working Memory			Math		Working Memory			Math
	VS	N	S	A	WPS	VS	N	S	A
Pretest									
Working Memory Visual Spatial (VS)									
Verbal-Numerals (N)	.48								
Verbal-Sentences (S)	.48	.50							
Math									
Arithmetic (A)	.45	.43	.40						
Word-Problem Solving (WPS)	.51	.43	.56	.58					
Posttest									
Working Memory Visuospatial (VS)	.65	.58	.53	.47	.48				
Verbal-Numerals (N)	.51	.44	.43	.40	.45	.57			
Verbal-Sentences (S)	.47	.54	.69	.42	.49	.55	.51		
Math									
Arithmetic (A)	.38	.34	.37	.80	.58	.40	.44	.41	
Word Problem Solving (WPS)	.52	.47	.53	.62	.73	.54	.49	.47	.66

[a] Correlations are reported for the subsample of students who received intervention: control group students with MD and classmates. This is because intervention was designed to disturb relations among WM and math performance.

All correlations are significant ($p < .001$). Visuospatial is *Automated Working Memory Assessment (AWMA;* Alloway, 2012*) Odd-One Out*; Verbal-Numerals is *WMTB-C (*Pickering & Gathercole, 2001*)–Counting Recall; Verbal Sentences* is *AWMA–Listening Recall; Arithmetic* is a sample of problems drawn from the school program, the district's online scope, and sequence and sample units for state standards, as is the case for *Second-Grade Word Problems*.

Table 5
Main Effects Multilevel Results from Figure 1 models (n=240)

Model/Parameter	Estimate	SE	*p*-value	*Cohen's d analog* ES
Outcome = post WM factor (see Figure 1a)				
Fixed effects, adjusted means:				
Intercept post VS_WM	11.008	0.699	<.0001	
Intercept post N_WM	13.934	0.44	<.0001	
Intercept post S_WM	6.81	0.508	<.0001	
Fixed effects of treatment (adj mean diff):				
GWM+P v. CON	**5.166**	**0.924**	**<.0001**	1.22
WPS+MWM v. CON	1.257	0.807	0.119	0.30
WPS v. CON	**2.621**	**0.759**	**0.001**	0.62
GWM+P v. WPS+MWM	**3.909**	**0.805**	**<.0001**	0.93
GWM+P v. WPS	**2.545**	**1.06**	**0.016**	0.60
WPS+MWM v. WPS	-1.364	0.93	0.142	-0.32
Factor loadings: Post WM factor by				
post VS_WM (loading)	1	--	--	
post N_WM (loading)	0.387	0.116	0.001	
post S_WM (loading)	0.168	0.132	0.203	
Fixed effects of pretest:				
Pretest VS_WM → Post VS_WM	0.262	0.07	<.0001	
Pretest N_WM → Post N_WM	0.237	0.093	0.011	
Pretest S_WM → Post S_WM	0.629	0.067	<.0001	
Variance Components: student-level				
post VS_WM res variance	10.946	4.933		
post N_WM res variance	12.286	1.804		
post S_WM res var	9.074	0.961		
post WM factor res variance	14.172	5.496		
Variance Components: class level				
post VS_WM int variance	2.059	1.100		
post N_WM int variance	1.553	1.686		
post S_WM int variance	3.6	1.081		
post VS_WM with post N_WM	1.456	0.783		
post VS_WM with post S_WM	1.689	1.591		
post N_WM with post S_WM	2.013	0.761		
Outcome = Posttest Arithmetic (see Figure 1b)				
Fixed effect, adjusted mean:				
Intercept	25.192	2.844	<.0001	
Fixed effects of treatment (adj mean diff):				
GWM+P v. CON	**8.231**	**1.21**	**<.0001**	0.64
WPS+MWM v. CON	**10.346**	**2.176**	**<.0001**	0.81
WPS v. CON	**9.182**	**2.496**	**<.0001**	0.71
GWM+P v. WPS+MWM	-2.115	2.271	0.352	-0.17
GWM+P v. WPS	-0.951	2.797	0.734	-0.07
WPS+MWM v. WPS	1.164	2.335	0.618	0.09
Fixed effects of pretest:				
Pretest arith → Post arith	1.13	0.116	<.0001	
Variance Components: student-level				
residual variance	98.668	10.612		
Variance Components: class level				
intercept variance	108.2	25.861		
Outcome = Delayed Posttest Arithmetic (see Figure 1b)				
Fixed effects, adjusted mean:				
Intercept	4.545	0.501	<.0001	
Fixed effects of treatment (adj mean diff):				
GWM+P v. CON	**1.265**	**0.29**	**<.0001**	0.57
WPS+MWM v. CON	**0.703**	**0.345**	**0.042**	0.32
WPS v. CON	**0.994**	**0.46**	**0.031**	0.45
GWM+P v. WPS+MWM	0.562	0.398	0.158	0.25
GWM+P v. WPS	0.271	0.423	0.522	0.12

WPS+MWM v. WPS	-0.291	0.437	0.506	-0.13
Fixed effect of pretest:				
Pretest arith → Delay-post arith	0.116	0.026	<.0001	
Variance Components: student-level				
residual variance	4.13	0.452		
Variance Components: class level				
intercept variance	1.668	0.624		
Outcome = Posttest WPS (see Figure 1b)				
Fixed effect, adjusted mean:				
Intercept	4.221	0.426	<.0001	
Fixed effects of treatment (adj mean diff):				
GWM+P v. CON	**1.806**	**0.424**	**<.0001**	0.23
WPS+MWM v. CON	**10.733**	**1.161**	**<.0001**	1.36
WPS v. CON	**11.868**	**1.397**	**<.0001**	1.51
GWM+P v. WPS+MWM	**-8.928**	**1.248**	**<.0001**	-1.13
GWM+P v. WPS	**-10.062**	**1.423**	**<.0001**	-1.28
WPS+MWM v. WPS	-1.135	2.019	0.574	-0.14
Fixed effect of pretest:				
Pretest WPS → Post WPS	0.637	0.316	0.044	
Variance Components: student-level				
residual variance	33.809	3.181		
Variance Components: class level				
intercept variance	2.612	1.672		
Outcome = Delayed Posttest WPS (see Figure 1b)				
Fixed effect, adjusted mean:				
Intercept	2.284	0.257	<.0001	
Fixed effects of treatment (adj mean diff):				
GWM+P v. CON	**1.008**	**0.276**	**<.0001**	0.29
WPS+MWM v. CON	**3.586**	**0.528**	**<.0001**	0.99
WPS v. CON	**4.51**	**0.698**	**<.0001**	1.24
GWM+P v. WPS+MWM	**-2.579**	**0.61**	**<.0001**	-0.71
GWM+P v. WPS	**-3.502**	**0.746**	**<.0001**	-0.97
WPS+MWM v. WPS	-0.923	0.921	0.316	-0.25
Fixed effect of pretest:				
Pretest WPS→ Delay Post WPS	0.28	0.145	0.054	
Variance Components: student-level				
residual variance	9.649	1.221		
Variance Components: class level				
intercept variance	0.925	1.069		

Table 6

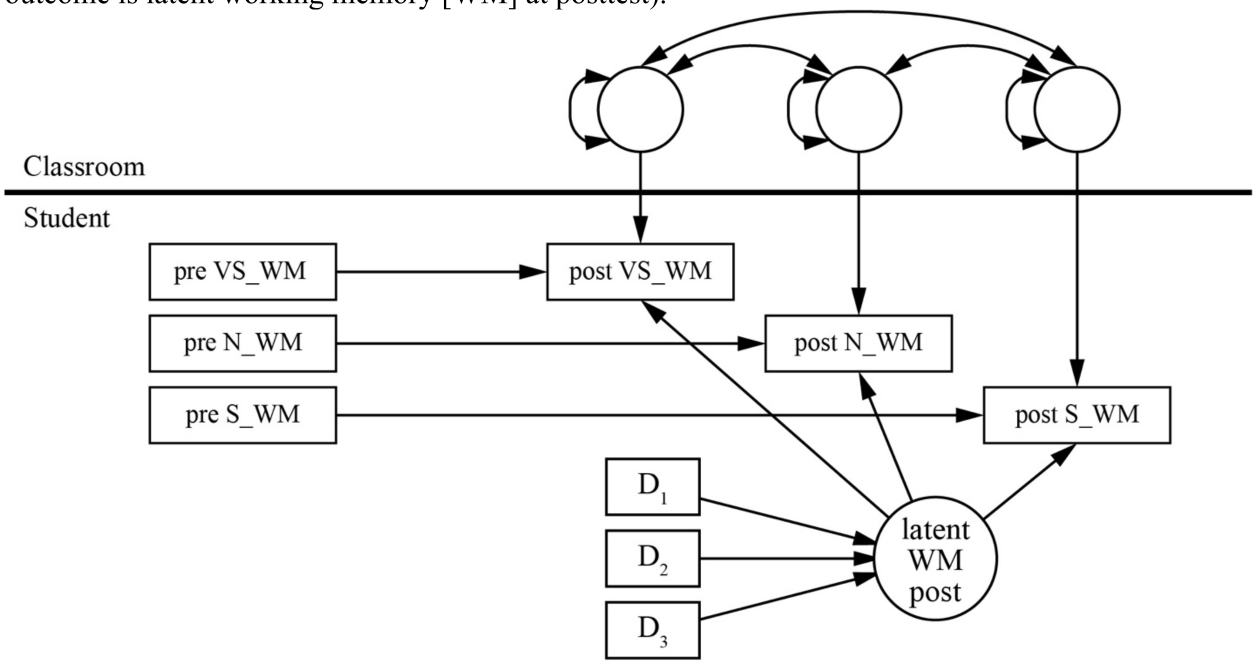
Indirect, Direct, a-Path, and b-Path Effect Estimates from the Multilevel Mediation Model in Figure 2

Model/Parameter	Estimate	SE	*p*-value
Outcome=Delayed posttest arithmetic, Mediator=WM factor, X=(GWM+P) v. CON			
"a-path": ([GWM+P] v. CON) → mediator	4.710	1.207	0.000
"b-path": mediator → outcome†	0.190	0.135	0.158
"direct effect": ([GWM+P] v. CON) → outcome	0.433	0.446	0.332
"indirect effect"= "a-path"×"b-path"	0.895 95%CI=(-.454, 1.817)		
Outcome=Delayed posttest arithmetic, Mediator=WM factor, X=(WPS+MWM v. CON)			
"a-path": ([WPS+MWM] v. CON) → mediator	1.299	0.745	0.081
"b-path": mediator → outcome†	0.190	0.135	0.158
"direct effect": ([WPS+MWM] v. CON) → outcome	0.442	0.410	0.281
"indirect effect"= "a-path"×"b-path"	0.247 95%CI=(-.144, .755)		
Outcome=Delayed posttest arithmetic, Mediator=WM factor, X=(WPS v. CON)			
"a-path": (WPS v. CON) → mediator	2.491	0.749	0.001
"b-path": mediator → outcome†	0.190	0.135	0.158
"direct effect": (WPS v. CON) → outcome	0.527	0.514	0.305
"indirect effect"= "a-path"×"b-path"	0.473 95%CI=(-.166, 1.363)		
Outcome=Delayed posttest WPS, Mediator=WM factor, X=([GWM+P] v. CON)			
"a-path": ([GWM+P] v. CON) → mediator	5.020	1.025	<.0001
"b-path": mediator → outcome†	0.203	0.099	0.041
"direct effect": ([GWM+P] v. CON) → outcome	-0.003	0.562	0.996
"indirect effect"= "a-path"×"b-path"	**1.019*** 95% CI=(0.048, 1.792)		
Outcome=Delayed posttest WPS, Mediator=WM factor, X=([WPS+MWM] v. CON)			
"a-path": ([WPS+MWM] v. CON)→ mediator	1.338	0.774	0.084
"b-path": mediator → outcome†	0.203	0.099	0.041
"direct effect": ([WPS+MWM] v. CON)→ outcome	3.264	0.567	<.0001
"indirect effect"= "a-path"×"b-path"	0.272 95%CI=(-.060, .745)		
Outcome=Delayed posttest WPS, Mediator=WM factor, X=(WPS v. CON)			
"a-path": (WPS v. CON) → mediator	2.623	0.730	<.0001
"b-path": mediator → outcome†	0.203	0.099	0.041
"direct effect": (WPS v. CON) → outcome	3.961	0.654	<.0001
"indirect effect"= "a-path"×"b-path"	**0.532*** 95%CI= (.018, 1.238)		
Outcome=Delayed posttest WPS, Mediator=WM factor, X=([GWM+P] v. [WPS+MWM])			
"a-path": ([GWM+P] v. [WPS+MWM])→ mediator	3.681	0.846	<.0001
"b-path": mediator → outcome†	0.203	0.099	0.041
"direct effect":([GWM+P] v. [WPS+MWM])→outcome	-3.267	0.710	<.0001
"indirect effect"= "a-path"×"b-path"	**0.747*** 95%CI=(.037, 1.318)		
Outcome=Delayed posttest WPS, Mediator=WM factor, X=([GWM+P] v. WPS)			
"a-path":([GWM+P] v. WPS)→ mediator	2.397	1.114	0.031
"b-path": mediator → outcome†	0.203	0.099	0.041
"direct effect":([GWM+P] v. WPS)→ outcome	-3.965	0.787	<.0001
"indirect effect"= "a-path"×"b-path"	0.487 95% CI=(-.052, .963)		

Notes. For each mediation pathway the total effect (*c*-path) can be computed as the direct effect plus the indirect effect. † = The same *b*-path pertains to all mediation tests involving the same mediator and same outcome. Though only estimated and tested once, a given *b*-path is repeatedly presented across rows of this table to underscore its role in testing different mediation pathways.

Figure 1. *Path diagrams of main effects multilevel models*.

Panel 1a: Main effects multilevel structural equation model with latent outcome (*latent WM post* outcome is latent working memory [WM] at posttest).



Panel 1b: Main effects multilevel model with manifest outcome (manifest outcome *post Y* is posttest arithmetic or delayed-posttest arithmetic or posttest word-problem solving [WPS], or delayed-posttest WPS).

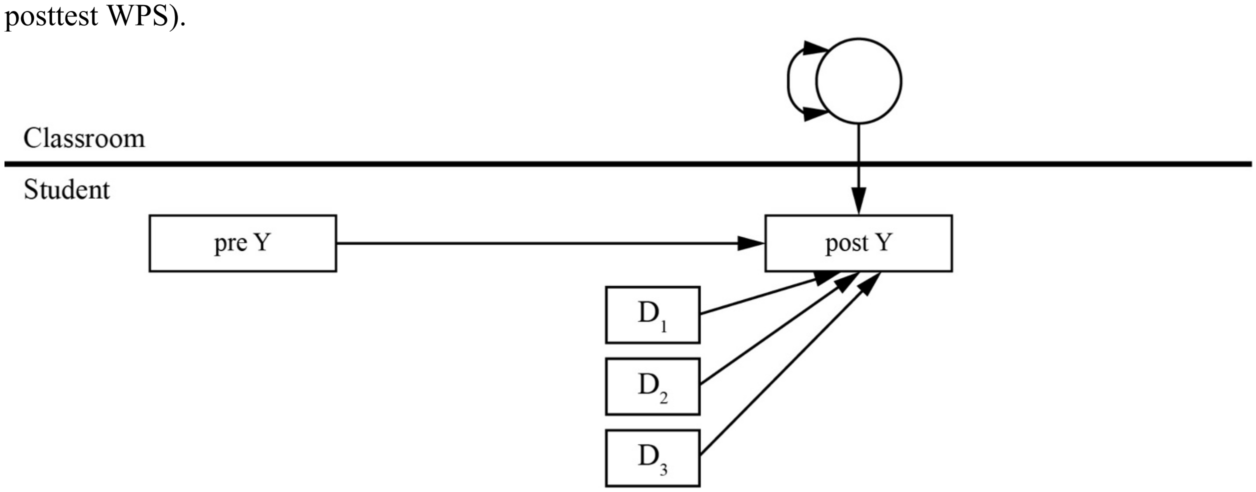
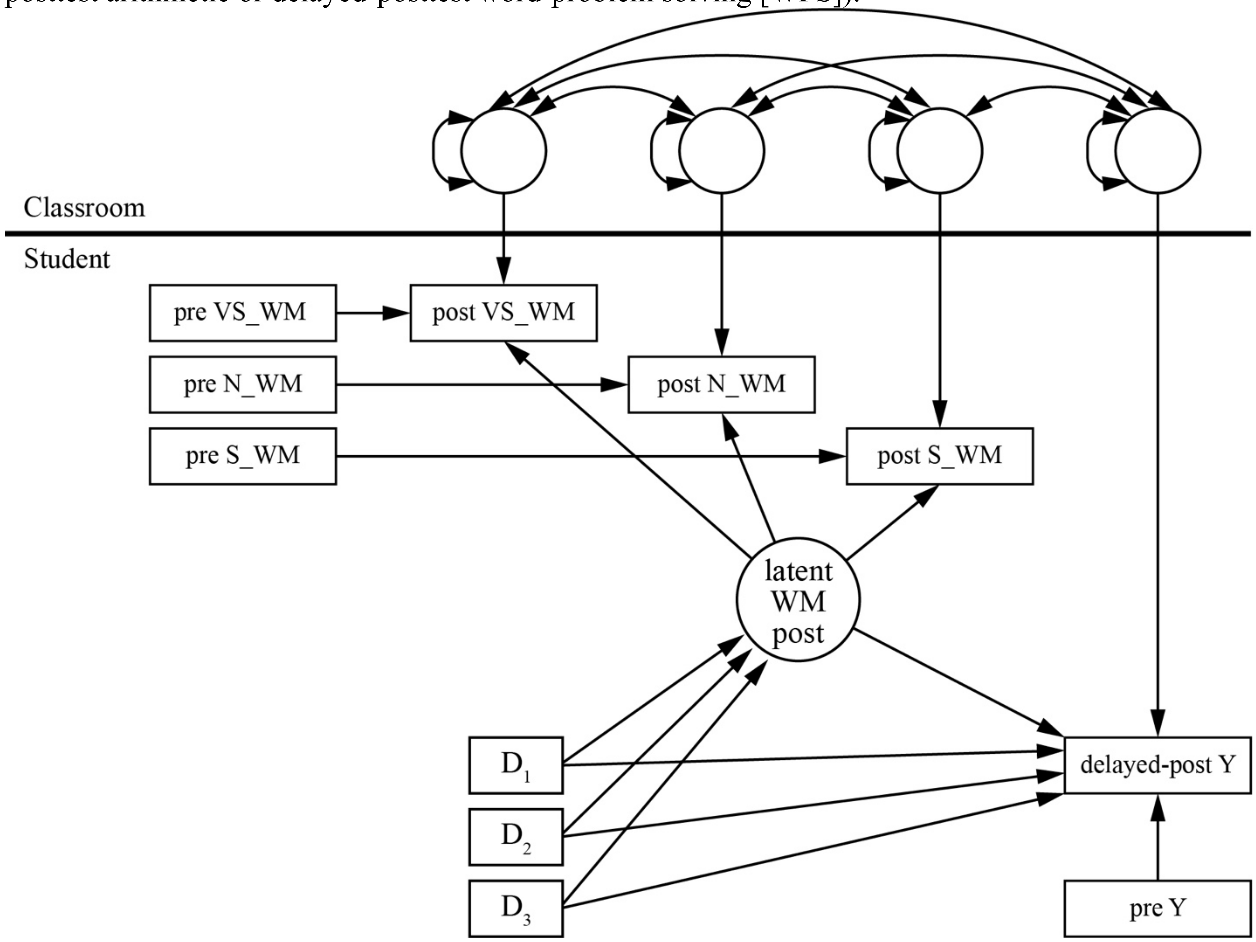
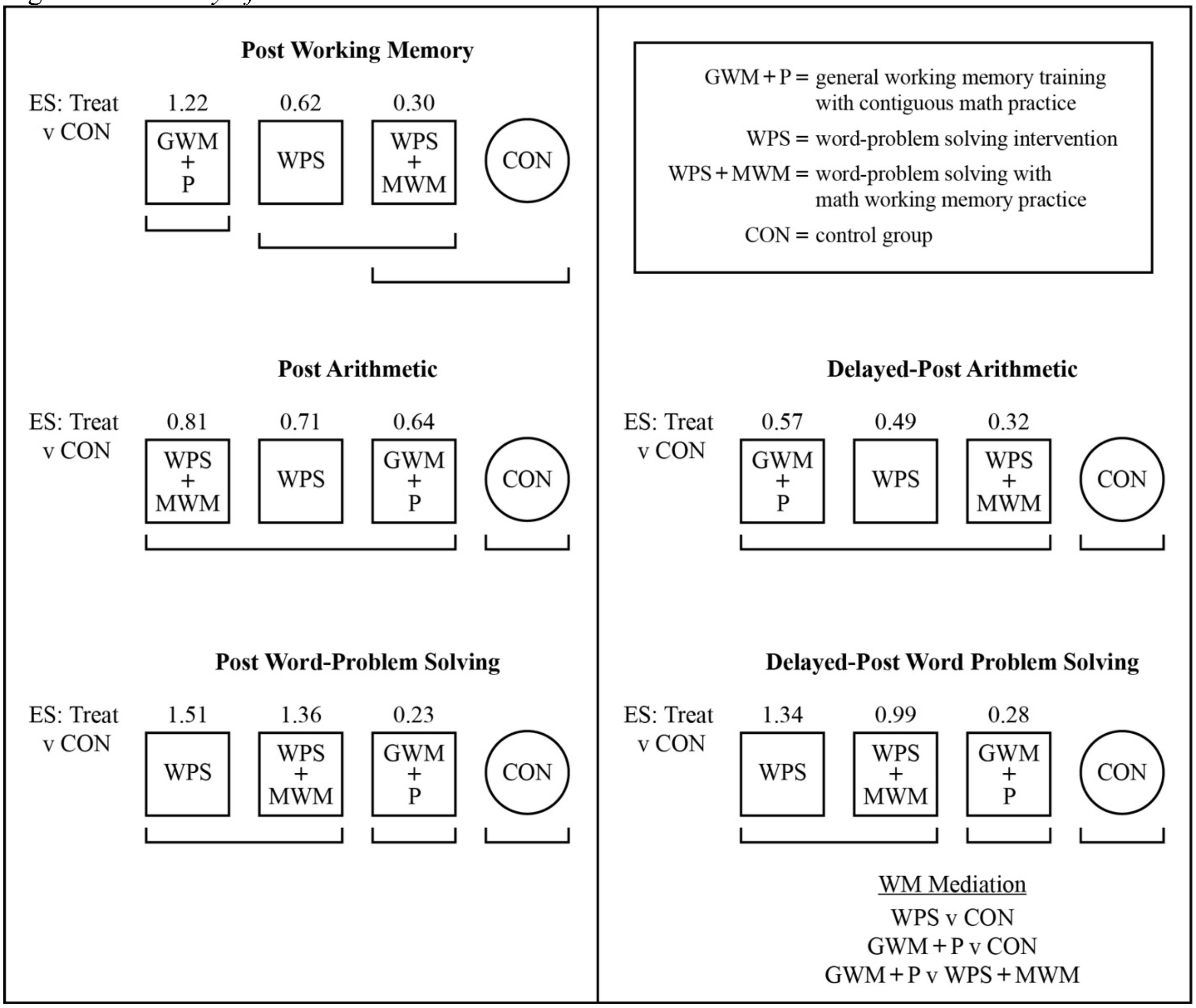


Figure 3. Path diagram of the multilevel mediation model (where *delayed-posttest Y* is delayed-posttest arithmetic or delayed-posttest word-problem solving [WPS]).



Notes: Curved arrows denote (co-)variances. Straight arrows denote regression paths. Circles denote latent variable. Rectangles denote manifest variables. D_1, D_2, and D_3 are dummy variables for treatment contrasts, as defined in the data analysis section. VS_WM (Visuospatial working memory) is *Automated Working Memory Assessment (AWMA;* Alloway, 2012*)– Odd-One Out*; N_WM (Verbal-*Numerals* working memory) is *WMTB-C (*Pickering & Gathercole, 2001*)–Counting Recall;* S_WM (Verbal *Sentences* working memory) is *AWMA–Listening Recall.*

Figure 3. *Summary of Results*



Notes: Conditions are ordered from largest to smallest effect size (ES). Bracketed lines under boxes indicate which conditions are comparable and which are significantly different. There were three significant WM mediation effects, all for the delayed-post word-problem solving outcome.