Journal of Educational Psychology
2014, Vol. 106, No. 2, 499–514

0022-0663/14/$12.00 DOI: 10.1037/a0034341

Does Working Memory Moderate the Effects of Fraction Intervention? An Aptitude–Treatment Interaction

Lynn S. Fuchs, Robin F. Schumacher,
Sonya K. Sterba, Jessica Long, Jessica Namkung,
Amelia Malone, and Carol L. Hamlett
Vanderbilt University

Nancy C. Jordan
University of Delaware

Russell Gersten
Instructional Research Group, Long Beach, California

Robert S. Siegler
Carnegie Mellon University

Paul Changas
Metropolitan-Nashville Public Schools, Nashville, Tennessee

This study investigated whether individual differences in working memory (WM) moderate effects of 2 variations of intervention designed to improve at-risk 4th graders' fraction knowledge. We also examined the effects of each intervention condition against a business-as-usual control group and assessed whether children's measurement interpretation of fractions mediated those effects. At-risk students (n = 243) were randomly assigned to control and 2 intervention conditions. The interventions each lasted 12 weeks, with three 30-min sessions per week. The major focus of both intervention conditions was the measurement interpretation of fractions. Across the 2 conditions, only 5 min of each 30-min session differed. One condition completed activities to build fluency with 4 measurement interpretation topics; in the other, activities were completed to consolidate understanding on the same 4 topics. Results revealed a significant aptitude–treatment interaction, in which students with very weak WM learned better with conceptual activities but children with more adequate (but still low) WM learned better with fluency activities. Both intervention conditions outperformed the control group on all outcomes, and improvement in the measurement interpretation of fractions mediated those effects.

Keywords: fractions, intervention, working memory, aptitude treatment intervention

Individual differences in children's cognitive resources are associated with mathematics learning, even when individual differences in foundational mathematics knowledge are statistically controlled (e.g., Fuchs, Geary, Compton, Fuchs, Hamlett, & Bryant, 2010; Fuchs, Geary, Compton, Fuchs, Hamlett, Seethaler, et al., 2010; Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Krajewski & Schneider, 2009; Kroesbergen, van Luit, Van Lieshout, Van Loosbroek, & Van de Rijt, 2009). This indicates that mathematics intervention should be designed to help students with poor foundational mathematics skills compensate for limitations in the cognitive resources associated with poor learning.

Geary et al. (2008) hypothesized that building automaticity with skills or strategies central to a new domain of learning enhances outcomes by compensating for such limitations. This fluency practice compensation hypothesis predicts that the benefits of fluency practice should be stronger for students with weaker cognitive abilities than for students with stronger abilities. Few studies, however, have examined whether or how children's cognitive abilities moderate (i.e., interact with) the effects of intervention. Identifying such aptitude–treatment interactions may provide the basis for "personalizing" intervention in ways that capitalize on children's cognitive strengths while reducing the negative consequences of limitations in those cognitive abilities.

In the present study, we tested the fluency practice compensation hypothesis in the context of fraction learning at fourth grade by assessing whether fluency practice differentially benefits stu-

This article was published Online First September 9, 2013.

Lynn S. Fuchs, Robin F. Schumacher, Sonya K. Sterba, Jessica Long, Jessica Namkung, Amelia Malone, and Carol L. Hamlett, Department of Special Education, Vanderbilt University; Nancy C. Jordan, School of Education, University of Delaware; Russell Gersten, Instructional Research Group, Long Beach, California; Robert S. Siegler, Department of Psychology, Carnegie Mellon University; Paul Changas, Department of Research, Assessment, and Evaluation, Metropolitan-Nashville Public Schools, Nashville, Tennessee.

This research was supported in part by Grant R324C100004 from the Institute of Education Sciences in the U.S. Department of Education to the University of Delaware, with a subcontract to Vanderbilt University, and by Core Grant HD15052 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development to Vanderbilt University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Institute of Education Sciences, the U.S. Department of Education, the Eunice Kennedy Shriver National Institute of Child Health and Human Development, or the National Institutes of Health.

Correspondence concerning this article should be addressed to Lynn S. Fuchs, 228 Peabody, Vanderbilt University, Nashville, TN 37203. E-mail: lynn.fuchs@vanderbilt.edu

dents with weaker working memory (WM). To provide context for this present study, we discuss the importance of fraction knowledge for success in and out of school. We also explain how the measurement interpretation of fractions, a major focus of intervention and major outcome of interest in the present study, has been shown to mediate overall fraction learning. Then we discuss the role of cognitive resources in the development of fraction knowledge; summarize prior work on the fluency practice compensation hypothesis and moderators of fraction intervention; and present the study hypotheses.

Importance of Fractions

Fraction knowledge is considered foundational for learning algebra, for success with more advanced mathematics, and for competing successfully in the American workforce (Geary, Hoard, Nugent, & Bailey, 2012; National Mathematics Advisory Panel [NMAP]; Siegler et al., 2013). Yet, half of middle and high school students in the United States are still not proficient with the ideas and procedures taught about fractions in the elementary grades (e.g., Behr, Harel, Post, & Lesh, 1992; National Council of Teachers of Mathematics [NCTM], 2007; NMAP; Ni, 2001). This prompted NMAP to recommend that high priority be assigned to improving fraction performance, especially understanding of fractions, given its importance for learning and maintaining accurate fraction procedures (e.g., Hecht, Close, & Santisi, 2003; Mazzocco & Devlin, 2008; Ni & Zhou, 2005; Rittle-Johnson, Siegler, & Alibali, 2001).

At fourth grade, this includes two types of understanding. The first is an understanding of the part–whole relationship, with which children understand a fraction as one or more parts of a single object (e.g., two or five parts of a pizza) or a subset of a group of objects (e.g., two of five pizzas). This type of understanding is largely intuitive, based on children's experiences with sharing, and is manifested as early as 4 years of age (e.g., Mix, Levine, & Huttenlocher, 1999). Part–whole understanding is typically represented using an area model, in which a region of a shape is shaded or a subset of objects is distinguished from other objects. This is the dominant approach in American schooling.

The second type of understanding is the measurement interpretation of fractions, which reflects cardinal size (Hecht, 1998; Hecht et al., 2003) and is often represented with number lines (e.g., Siegler, Thompson, & Schneider, 2011).[1] The measurement interpretation is less intuitive than part–whole understanding. Although it can be linked to children's experiences with measuring, it depends far more on formal instruction that explicates conventions of symbolic notation, the inversion property of fractions (i.e., fractions with the same numerator become smaller as denominators increase), and the infinite density of fractions on any given segment of the number line. Measurement interpretations of both fractions and whole numbers are related to a variety of cognitive processes, especially WM and executive functioning (Bull & Scerif, 2001; Passolunghi & Pazzaglia, 2004, 2005; Siegler & Pyke, 2012). Improvement in the measurement interpretation of fractions is thought to be a key mechanism in explaining the development of fraction knowledge (Geary et al., 2008), and Fuchs, Schumacher, et al. (2013) found support for this idea.

Accordingly, in the present study, we randomly assigned at-risk children to three conditions, two of which focused mainly on the measurement interpretation of fractions. We contrasted the effects of these two intervention conditions against a business-as-usual condition, which controlled for maturation and for the effects of the typical school program that focused mainly on part–whole understanding.

Of greater interest, however, was the distinction between the two 36-session intervention conditions. In all 30-min sessions, 25 min were identical across both conditions. Only one 5-min segment, which provided supplementary activities, differed. In the fluency condition, students completed strategic speeded activities to build fluency on four measurement interpretation topics. In the conceptual condition, students explained their reasoning about the same four measurement interpretation topics, with the aid of manipulatives. Essentially, the fluency condition was designed to help student automatize the procedural steps for accurate solution on the four measurement topics, whereas the conceptual condition was designed to consolidate the meaning behind the procedures represented in those topics (see Rittle-Johnson & Alibali, 1999; Tall et al., 2001). We note, however, that the larger intervention, which contextualized both conditions, focused on understanding of the ideas represented in fractions as well as procedural strategies. We contrasted the effects of these two conditions against each other and against the control group on the measurement interpretation of fractions, procedural calculation skill (adding/subtracting fractions), and released fraction items from the National Assessment of Educational Progress (NAEP). Our major research question, however, was whether effects of the fluency versus conceptual conditions differed as a function of students' WM.

Does WM Moderate the Effects of Intervention?

This question speaks to Geary et al.'s (2008) *fluency practice compensation hypothesis*: that fluency practice especially enhances the learning of at-risk students by building automaticity with foundational skills, because it compensates for those children's limitations in the cognitive resources associated with development. Among the resources thought to affect mathematics learning, WM features prominently (Geary, 2004). (We define *WM* as a resource-limited capacity that allocates attention and plans, sequences, and maintains information in short-term memory while processing the same or other information; Baddeley & Logie, 1999; Unsworth & Engle, 2007).

Studies show that individual differences in WM contribute to performance on fraction tasks (Fuchs, Schumacher, et al., 2013; Hecht et al., 2003; Hecht & Vagi, 2010, but also see Seethaler, Fuchs, Star, & Bryant, 2011, and Jordan et al., 2013), even when controlling for other cognitive and linguistic variables and prior mathematics achievement. Yet, debate exists about whether individual differences in WM are driven by more fundamental differences in processing speed or whether the attentional focus associated with the central executive speeds information processing. The measurement interpretation of fractions may especially tax such systems, because it requires children to simultaneously consider the contribution of numerators and denominators when completing

[1] The measurement interpretation of fractions involves more than number lines. It can be represented in other ways, such as measuring cups. It also involves other tasks such as comparing or ordering fraction magnitudes.

tasks such as comparing fractions or placing them on a number line. And at-risk children often experience limited WM capacity, inattentive behavior, and slow processing speed (e.g., Compton, Fuchs, Fuchs, Lambert, & Hamlett, 2012).

This suggests that reducing cognitive load could improve the efficacy of interventions that stress the measurement interpretation of fractions. Reductions could be accomplished by (a) building fluency with efficient strategies for chunking (recoding a multidimensional concept into fewer dimensions) or segmenting (breaking a task into a series of steps, each of which is less resource demanding), and (b) automatizing retrieval of fractional values in relation to benchmark fractions, such as one half (Fuchs, Schumacher, et al., 2013). In our fluency practice intervention, we adopted such methods, with the goal of reducing demands on (i.e., compensating for poor) WM, processing speed, and related attentional processes.

We therefore hypothesized that individual differences in these cognitive abilities would moderate the effects of intervention. With such moderation, the relation between these cognitive abilities and fraction learning would be expected to differ as a function of condition. Because fluency practice was designed to compensate for such limitations, we expected students in the fluency condition to respond similarly well regardless of cognitive abilities but expected that students in the conceptual condition with lower abilities would experience poor outcomes relative to students in the conceptual condition with stronger abilities. We focused on WM but controlled for attentive behavior and processing speed, as well as language ability (because our measure of WM relied on language), in the statistical model.

We identified two relevant prior studies. Fuchs, Geary, et al. (2013) examined the effects of a number knowledge intervention with fluency versus conceptual activities on at-risk first graders' emerging competence with whole-number arithmetic and explored whether reasoning ability, a cognitive resource associated with whole-number arithmetic (e.g., Geary et al., 2012), moderated the effects of study condition. At-risk children with low reasoning ability who participated in the conceptual condition suffered poor outcomes relative to peers with medium and high reasoning ability in the same condition. By contrast, the effects of fluency practice were strong regardless of children's reasoning ability. So, in line with the fluency practice compensation hypothesis, fluency practice helped at-risk students compensate for weak reasoning ability. By contrast, the conceptual condition, with its sole focus on number relations and principles, demanded greater reasoning ability in ways that limited outcomes for the subset of learners with poor reasoning ability.

The other prior study (Fuchs, Schumacher, et al., 2013) focused on fractions. It was not designed to test the fluency practice compensation hypothesis (i.e., the study did not contrast intervention conditions with vs. without fluency practice), but it did examine cognitive moderators of the effects of the intervention versus the control group. For some outcomes and cognitive processes, the relation between students' cognitive abilities and their learning was significantly stronger in the control than in the intervention condition. That is, response in the intervention condition, designed to compensate for limitations in cognitive abilities, depended less or not at all on the children's cognitive abilities. But for other outcomes and cognitive processes, intervention children with stronger cognitive resources (including WM, attentive behavior, and processing speed) learned better than intervention children with weaker performance on these cognitive resources.

Study Hypotheses

We extended the literature by testing the fluency practice compensation hypothesis in the context of fourth-grade fraction learning. We had three hypotheses. The first two were that both intervention conditions would promote stronger learning than the control group and that improvement in children's measurement interpretation would mediate effects between the intervention versus control conditions. Such results provide the basis for testing the third hypothesis: that WM would moderate the effects of fluency versus conceptual supplementary activities, such that the benefits of fluency practice would be stronger for students with weaker WM than for students with stronger WM (while controlling for processing speed, attentive behavior, and language ability). In this way, we expected the relation between children's WM and their fraction learning to differ as a function of treatment condition. To examine differential relations, we included at-risk students with a range of performance on the cognitive abilities across the three conditions.

Method

Overview

To review, at fourth grade, we contrasted the effects of two intervention conditions, both of which (a) occurred over 12 weeks, three times per week, 30 min per session; (b) focused mainly on the measurement interpretation of fractions; and (c) incorporated the same instructional design. Supplementary activities, conducted for 5 min of each session, differed between the two conditions. One condition provided strategic speeded activities designed to build fluency on four measurement interpretation topics. The other was designed to consolidate understanding by having students explain their reasoning about the same measurement interpretation topics with the aid of manipulatives. We also included a control group that represented business-as-usual instruction, focused more on part–whole understanding than either of the two intervention conditions. Before intervention began, we assessed children's WM, processing speed, attentive behavior, and language. Before and after intervention, we assessed their fraction knowledge using a well-accepted measure of the measurement interpretation of fractions, a measure of procedural calculation skill with fractions, and released items from the NAEP, an external, widely accepted measure of fraction competence. To contextualize results, we compared at-risk students' year-end performance on group-administered fraction measures against the performance of low-risk classmates.

Participants

We defined risk as performance below the 35th percentile on a broad-based calculations assessment (Wide Range Achievement Test–4 [WRAT]; Wilkinson & Robertson, 2006). To ensure strong representation across the range of scores below the 35th percentile, we sampled half the at-risk (AR) students from below the 15th percentile; the other half was sampled from between the 15th and

34th percentiles.[2] Because this study was not about intellectual disability, we administered the two-subtest Wechsler Abbreviated Scales of Intelligence (WASI; Wechsler, 1999) to all students who met the risk criterion, and we excluded 18 children with T scores below the 9th percentile on both subtests. We sampled between two and eight AR students per classroom, stratifying by more versus less severe risk in each classroom. After exclusion and random sampling, the sample comprised 277 AR students from 49 classrooms in 14 schools. We randomly assigned these students at the individual level, stratifying by classroom and risk severity, to fluency activities ($n = 94$), conceptual activities ($n = 91$), and the control condition ($n = 92$). Another 281 low-risk classmates (>34th percentile) were randomly sampled to represent each of the 49 classrooms in similar proportion to AR students from those classrooms. These low-risk classmates served as a comparison group.

Of the 277 AR students, 34 moved (10 fluency; 12 conceptual; 12 control) before the end of the study. These students did not differ statistically from the remaining AR students on pretest measures and did not differ significantly on any pretest measure as a function of condition. We omitted these children, leaving 243 students in the final AR sample: 84 fluency practice; 79 conceptual practice; and 80 control. Among the 281 low-risk students, 16 students who moved did not differ from remaining students on pretest measures. We omitted them, leaving 265 in the low-risk comparison group.

On the screening measure, low-risk students performed reliably higher than each of the AR groups, which performed comparably. The mean WRAT standard score was 85.46 ($SD = 6.29$) for fluency, 85.52 ($SD = 6.74$) for conceptual, 84.44 ($SD = 7.10$) for control, and 104.23 ($SD = 6.85$) for low-risk classmates. There were no significant differences among the AR conditions on WASI IQ, with mean standard scores of 94.25 ($SD = 12.20$) for fluency, 93.70 ($SD = 11.98$) for conceptual, and 93.30 ($SD = 11.57$) for control. In the fluency, conceptual, and control groups, respectively, the percentage of girls was 63, 62, and 59; the percentage of English learners was 14, 14, and 5; the percentage receiving subsidized lunch was 93, 95, and 86; the percentage receiving special education was 8, 10, and 12. In the fluency condition, the percentages of African American, White, Hispanic, and other students was 58, 17, 24, and 1; in conceptual condition, 61, 14, 22, and 3; and in control, 58, 16, 22, and 4 (all Hispanic students were White). Thus, the AR groups were demographically comparable (all $ps > .05$). We did not collect demographic data (or individually administered measures) on low-risk students.[3]

Screening Measures

The math screening measure was *WRAT-4–Arithmetic* (Wilkinson & Robertson, 2006), with which students complete calculation problems of increasing difficulty. Alpha on this sample was .85. The IQ screening measure was the Wechsler Abbreviated Scales of Intelligence (Wechsler, 1999). *Vocabulary* assesses expressive vocabulary, verbal knowledge, memory, learning ability, and crystallized and general intelligence; subjects identify pictures and define words. *Matrix Reasoning* measures nonverbal fluid reasoning and general intelligence; subjects select 1 of 5 options that best completes a visual pattern. Reliability exceeds .92.

Moderator Effect Measures

WM. To assess the central executive component of WM, we used the Listening Recall subtest of the *Working Memory Test Battery for Children* (WMTB-C; Pickering & Gathercole, 2001). The task includes six dual-task items at span levels from 1–6 to 1–9. Passing four items at a level moves the child to the next level. At each span level, the number of items to be remembered increases by one. Failing three items terminates the subtest. For each item, the child determines if each sentence is true; then recalls the last word in the series of sentences. We used the trials correct score. Test–retest reliability ranges from .84 to .93.

Attentive behavior. The Strength and Weaknesses of ADHD Symptoms and Normal Behavior scale (*SWAN*; Swanson et al., 2004) samples items from the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; American Psychiatric Association, 1994) criteria for attention-deficit/hyperactivity disorder for inattention (9 items) and hyperactivity-impulsivity (9 items), but scores are normally distributed. Teachers rate items on a 1–7 scale. We report data only for the inattentive subscale, as the average rating across the nine items. The SWAN correlates well with other dimensional assessments of behavior related to attention (www.adhd.net). Alpha for the inattentive subscale on the present sample was .96.

Processing speed. *Cross Out* from Woodcock-Johnson III (Woodcock, McGrew, & Mather, 2001) measures processing speed by asking students to locate and circle five pictures that match a target picture in that row; students have 3 min to complete 30 rows. Reliability is .91.

Language. *WASI Vocabulary* (Wechsler, 1999) measures expressive vocabulary, verbal knowledge, and foundation of information with 42 items. The first four items present pictures; the student identifies the object in the picture. For the remaining items, the tester says a word that the student defines. Responses are awarded a score 0, 1, or 2 depending on quality. Testing is discontinued after five consecutive scores of 0. The score is the total number of points. As per Zhu (1999), split-half reliability is .86–.87; the correlation with the Wechsler Intelligence Scale for Children (3rd ed.) Full Scale IQ is .72.

Fraction Measures

Measures, although similar to those in Fuchs, Schumacher, et al. (2013), were revised to reflect the increased difficulty of the scope of intervention in the present study.

[2] To increase the possibility of identifying comparable numbers of students with more and less severe risk in the same classrooms, we sampled AR students in each classroom for participation within severity bands. We then stratified by risk severity when randomly assigning students to the three study conditions. In some classrooms, we had numbers of more versus less severe risk that permitted an equal numbers of students with more versus less severe risk to be randomly assigned to the three conditions. In other classrooms, the number of participants required one or more extra students with more or less severe risk in one or another condition.

[3] Low-risk classmates were only pre- and posttested on group-administered measures for two reasons. First, because the study was not about low-risk students, we did not think it appropriate to spend school time on the individual assessment battery or to ask teachers to spend time completing demographic forms or attentive behavior ratings on these students. Second, study resources did not permit us to administer the individual test battery to an additional 281 children.

To assess the measurement interpretation of fractions, we used the *Fraction Number Line* (Hamlett, Schumacher, & Fuchs, 2011, adapted from Siegler et al., 2011), which requires students to place proper fractions, improper fractions, and mixed numbers on a number line labeled with endpoints of 0 and 2. For each trial, the number line is presented, and a target fraction is shown in a large font below the line. Students practice with two target fractions and then estimate the location of 20 test items: 12/13, 7/9, 5/6, 1/4, 2/3, 1/2, 1/19, 3/8, 7/4, 3/2, 4/3, 7/6, 15/8, 1 1/8, 1 1/5, 1 5/6, 1 2/4, 1 11/12, 5/5, and 1. Items are presented in random order. The score for each item is the absolute difference between the child's placement and the correct position of the number. Scores are averaged across the 20 items and divided by 2 (the numerical range of the number line). When multiplied by 100, scores are equivalent to the percent of absolute error (PAE), as reported in the literature. Lower scores indicate stronger performance (but note that in some analyses, we multiplied scores by −1). Test–retest reliability, on a sample of 63 students across 2 weeks, was .80.

To index generalized learning about fractions, we administered 19 released fraction items from 1990–2009 *NAEP* mathematics released items: easy, medium, or hard from the fourth-grade assessment or easy from the eighth-grade assessment. Testers read each problem aloud (with up to one rereading upon student request). Eight items assess the part–whole interpretation (e.g., provided with a rectangle divided into six equal parts, the student is directed to shade 1/3); nine assess measurement interpretation (e.g., provided with four lists of three fractions, the student is asked, "In which of the following are the three fractions arranged from least to greatest?"; one requires subtraction with like denominators; and one asks how many fourths make a whole. (Of the nine measurement interpretation items, two involve a number line. We ran analyses with and without those two items included. Results were not meaningfully different, so we conducted analyses on the full set of items.) Students select an answer from four choices (11 problems); write an answer (3 problems); shade a portion of a fraction (1 problem); mark a number line (2 problems); write a short explanation (1 problem); or write numbers, shade fractions, and explain the answer (1 problem with multiple parts). The maximum score is 25. Alpha on this sample was .81.

To index skill with fraction procedures, we administered two subtests from the Fraction Battery 2011—Revised (Schumacher, Namkung, Malone, & Fuchs, 2011). *Fraction Addition* includes five addition problems with like denominators and seven addition problems with unlike denominators. Six are presented vertically and six horizontally. *Fraction Subtraction* includes six subtraction problems with like denominators and six with unlike denominators; half are presented vertically and half horizontally. For each subtest, administration is terminated when all but two students have completed the test. Items are scored as 1 point for finding the correct numerical answer; 2 points if the item is appropriately reduced one time (7 items on addition and 8 items on subtraction); or 3 points if the item is appropriately reduced two times (1 subtraction item). We used the total score across these tests, which correlated .81. The maximum score across the tests is 41. Alpha on this sample was .93.

Control Group Instruction: Distinctions From the Intervention Conditions

Control group instruction relied on enVisionMATH (Pearson Education, 2011), the district's mathematics curriculum. It addresses fraction content at fourth grade in two units: Understanding Fractions and Adding and Subtracting Fractions, with 70% of lessons allocated to understanding fractions. For understanding fractions, the program relies mainly on part–whole understanding by using shaded regions and other manipulatives related to the area model, while encouraging students to write and draw when explaining fraction concepts. In a single lesson, benchmark fractions and equivalent fractions are introduced to teach students to make magnitude decisions (number lines are not included). Adding and Subtracting Fractions are taught via procedural rules.

The instructional design incorporates six elements. Essential Understanding presents new concepts and skills. The Daily Spiral Review provides practice on those concepts and skills and always includes a word problem. The third and fourth activities, Problem-Based Interactive Learning and Visual Learning, focus on conceptual understanding via interactive conversations and visual displays. In Assessment and Prescription, teachers check mastery of the daily objective to determine individual needs, and in Data-Driven Differentiated Instruction, teachers provide instruction in a leveled manner through assessments, homework, small-group center activities, and the schools' intervention period. According to What Works Clearinghouse (WWC; U.S. Department of Education, Institute of Education Sciences, 2013, p. 1), EnVisionMATH has "potentially positive effects." WWC reported an improvement index of six percentile points; the one study meeting WWC standards (Resendez & Azin, 2008) reported an average effect size of 0.15 at Grades 2 and 4. Also, most at-risk intervention students received intervention during the school's intervention period (typically 45 min). Intervention was delivered by certified teachers in groups that were smaller than the full class size but often larger than three students. It was provided in addition to the instruction that low-risk peers received. Fractions represented a major but not the exclusive focus.

In terms of content, there were three major distinctions between the control group and the two intervention conditions. First, the control group focused dominantly on part–whole understanding, whereas both intervention conditions emphasized the measurement interpretation of fractions. Second, the control group addressed some advanced skills not covered in the intervention conditions, such as estimation and word problems. Third, the control group did not restrict the range of fractions, whereas the intervention conditions limited the pool of denominators to 2, 3, 4, 5, 6, 8, 10, and 12 and the pool of equivalent fractions and reducing activities to 1/2, 1/3, 1/4, 1/5, and 1/1.

The Two Intervention Conditions: Commonalities

In all 30-min intervention sessions, 25 min were identical in the two intervention conditions. In this section, we describe the common program.

Delivery and training. This study's intervention was provided during one of three school periods (depending on teachers' scheduling preferences): during part of the math block (typically 50 min) or math center time (typically 20 min) or the school's intervention period (typically 45 min), such that the amount of mathematics instructional time was similar for AR intervention and control students. In the fluency and conceptual practice conditions, intervention occurred in small groups (3:1). Tutors were full-time or part-time graduate-student employees of the research grant. Some were licensed teachers; most were not. Each was responsible for two to four groups, distributed over the fluency and conceptual practice conditions. We avoided contamination by color coding materials, regularly monitoring fidelity of implementation tapes (see below), and providing guidance in weekly meetings (see below).

Tutors were initially trained in a weeklong workshop, with follow-up trainings occurring biweekly for one hour. The additional trainings provided opportunities for (a) dynamic feedback as the fraction lessons progressed in difficulty and (b) problem solving in terms of students' challenging behavior and skill-level differences within groups. The intervention program, *Fraction Face-Off!* (Fuchs & Schumacher, 2011), was organized in a manual that included all materials and scripts for the 36 lessons. Fraction Face-Off! is a revision of *Fraction Challenge* (Fuchs, Schumacher, et al., 2013); the major difference is that Fraction Face-Off! addresses a more ambitious set of skills, aligned with the Common Core State Standards. The manual scripts provide a model for each lesson sequence and important explanatory language. Tutors reviewed but did not read from or memorize scripts. Also, prior to delivering lessons, they practiced delivering lessons with fellow tutors. This helped promote a high level of implementation fidelity while preserving teaching authenticity and responsiveness to student misunderstandings.

Content. The major focus in both intervention conditions was the measurement interpretation of fractions, which focused primarily on representing, comparing, ordering, and placing fractions on number lines. This focus was supplemented by attention to part–whole interpretation (e.g., showing objects with shaded regions) and fair shares representations to build on prior knowledge and classroom instruction. Number lines, fraction tiles, and fraction circles were used to explain concepts throughout the 36 lessons, with a greater emphasis on these visuals at the start of the program. We focused on proper fractions and improper fractions equal to one in Lessons 1–15 and introduced improper fractions > 1 and $<$ 2 in Lesson 16. We taught students to convert between improper fractions and mixed numbers; to place fractions on a 0–2 number line; and to order, compare, add, and subtract improper fractions and mixed numbers. We also included a focus on fraction calculations (i.e., addition and subtraction), but approximately 85% of content was allocated to concepts.

In Weeks 1–2, tutors addressed the meaning of a fraction, taught vocabulary (e.g., *numerator*, *denominator*, *unit, equivalent*), and identified shapes divided into equal parts. Tutors initially relied on a combination of part–whole relations, measurement, and equal sharing to explain fraction magnitude and then emphasized the roles of the numerator and denominator. Students practiced naming fractions, reading fractions, and comparing two fractions when denominators are the same, when numerators are the same, and when fractions equal a whole.

In Weeks 3–5, fractions equivalent to 1/2 2/4, 3/6, 4/8, 5/10, 6/12 were addressed, along with chunking/segmenting strategies for comparing two fractions in which the numerators and denominators both differed, using 1/2 as a benchmark for comparison and using $<$, $>$, and $=$ signs. Then, tutors introduced ordering three fractions and placing two fractions on the 0 to 1 number line marked with 1/2. Students continued to use 1/2 as a benchmark, as tutors also taught them to think about a fraction's relationship to 0 and 1 on the number line. (None of the number line activities relied on computers, as done in the pre/posttest assessment task.)

In Week 6, tutors introduced improper fractions and mixed numbers on the 0–2 number line and taught students to convert between improper fractions and mixed numbers. Tutors demonstrated the conversion with fraction circles while converting them symbolically, so students could visualize the mathematical procedure and build understanding. In Weeks 7–8, tutors integrated improper fractions and mixed numbers into the fraction comparing and ordering activities, while they cumulatively reviewed all concepts. Week 9 focused on simple fraction calculations, with ongoing review of previously introduced concepts. Tutors introduced addition and subtraction with like denominators (proper and improper fractions); addition and subtraction with unlike denominators (proper and improper fractions); and addition and subtraction with mixed numbers.

In Week 10, tutors increased difficulty by deleting the benchmark fraction, 1/2, from the 0–1 and 0–2 number lines and removing the whole number 1 from the 0–2 number line. They also focused on multiplicative reasoning by teaching children to find equivalent fractions via multiplication for 1/3, 1/4, and 1/5. Weeks 11 and 12, which were cumulative review, included the Fraction Championship, a game in which students competed by solving fraction problems of varying degrees of difficulty with predetermined point values.

Lesson activities. Each 30-min lesson comprised four activities, with activity names reflecting a sports theme. In the first activity (8–12 min), "Training," tutors introduced concepts, skills, problem-solving strategies, and procedures, while relying on manipulatives (e.g., fraction tiles, fraction circles, number lines) and visual representations. The "Relay" (8–12 min) involved group work on concepts and strategies taught during that day's Training. Students took turns completing problems while explaining their work to the group. All students simultaneously showed work for each problem on their own papers. The third activity (5 min), "Sprint" or "Conditioning" (depending on treatment condition), provided supplementary activities (see below). It was introduced in Week 3; in Weeks 1–2, Training or the Relay was extended to provide a total of 30 min. In the final activity, the "Individual Contest" (5 min), students independently completed paper-pencil problems on content representing that day's and previous Training topics.

Commonalities and Distinctions Between the Two Intervention Conditions

Commonalities. Training, the Relay, and the Individual Contest were identical across the two intervention conditions. The distinctions between the two intervention conditions pertained only to the third (5-min) activity, introduced in Week 3. Even within this third activity, however, the activities in the two conditions focused on identical content: for Topic 1 (Weeks 3–4), identifying fractions equal to 1/2 (e.g., 4/8, 2/6); for Topic 2 (Weeks 5–6), comparing two proper fractions with the same numerator or denominator (e.g., 1/3 vs. 1/8; 5/12 vs. 11/12); for

Topic 3 (Week 7), distinguishing among proper fractions, improper fractions, and mixed number (e.g., 7/8, 4/4, 5/3, 1 1/4); and for Topic 4 (Weeks 8–12), comparing fractions of increased difficulty (e.g., 1/2 vs. 5/10; 1/1 vs. 4/4; 5/6 vs. 7/4; 1 1/4 vs. 1 1/2; 7/4 vs. 7/6; 7/4 vs. 5/4).

Distinctions: Fluency condition. This third activity differed in terms of the activities students completed on these topics. In the 5-min fluency condition, students completed the Sprint. The group worked as a team to answer flashcards, trying to answer as many cards as possible within 2 min and attempting to meet or beat the previous lesson's score. Whenever an error occurred, students were required to provide a correct response, using a specified procedure, before the next flashcard was revealed. Time continued to elapse during the correction; this encouraged students to be careful but quick in their initial response. Each student typically had 4–8 opportunities to respond during the 2-min activity. The group score for each of the three sessions in the same week was graphed. The goal was to meet or beat the previous session's score during that week.

For Topic 1, each flashcard showed one fraction; students took turns stating "equal" or "not equal" to 1/2. The correction procedure required them to explain how the doubling rule (i.e., if you double the numerator and it equals the denominator, the fraction equals 1/2—a strategy explained in other parts of lessons) produces the correct response, as time continued to elapse. For Topics 2 and 4, each flashcard showed two fractions; students took turns stating which fraction is bigger. The correction procedure varied depending on the type of comparison. For comparisons between two fractions with the same denominator (e.g., 3/8 and 7/8), students explained, "7/8 is bigger because the same denominator means bigger numerator bigger fraction." For comparisons between proper and improper fractions (e.g., 1/2 and 4/3), students explained, "4/3 is bigger because it's improper and 1/2 is proper; improper fractions are equal to or greater than 1 and proper fractions are less than 1." For Topic 3, each flashcard showed one fraction; students took turns stating whether it was proper, improper, or mixed. The correction procedure required students to state the rule for identifying each fraction type (for proper fractions, "the numerator is smaller than the denominator so the fraction is less than 1 whole"; for improper fractions, "the numerator is the same or bigger than the denominator so the fraction is equal to or greater than 1 whole"; for mixed numbers, "a whole number and a proper fraction together makes a mixed number").

Distinctions: Conceptual practice. In the 5-min conceptual condition, students completed "Conditioning." Students in the group were given different fractions to show with fraction circles or fraction tiles; then they explained their thinking to the group. Students earned points for showing and explaining fractions. A game board showed the number of points required to meet the week's goal. At the end of each session in that week, they recorded the collective number of points earned. The tutor monitored the group's performance to identify a realistic but ambitious goal for the next week.

For Topic 1, each flashcard showed one fraction. Students represented it with fraction circles or tiles and stated if it was equal or not equal to 1/2. They earned one point for showing it correctly and one point for stating whether it was equal to 1/2. They corrected errors but did not earn points for errors; this was the case for all topics within Conditioning. For Topics 2 and 4, each flashcard showed two fractions. Students represented them with fraction circles or tiles and stated which was bigger. They earned one point for their explanation and one point per fraction they showed correctly. For Topic 3, each flashcard showed one fraction, which students represented with fraction circles or tiles. They stated if it was a proper fraction, improper fraction, or mixed number. (Students used the unit to represent the whole number 1 but not the unit in improper fractions; e.g., for 4/3, students used four 1/3 pieces, three of which were displayed as 1 whole with the remaining 1/3 next to it.)

Promoting Task-Oriented Behavior in Both Intervention Conditions

Because AR students often display attention, motivation, and self-regulation difficulties that may affect learning (e.g., Montague, 2007), we encouraged students to regulate their attention/behavior and to work hard. Tutors taught students that *on-task behavior* means listening carefully, working hard, and following directions and that on-task behavior is important for learning. Tutors set a timer to beep at three unpredictable times during each lesson, so students could not anticipate intervals. If all students were on task when the timer beeped, all students received a checkmark (if any one student was off task, no one earned a checkmark).

Students also earned check marks for correct work as follows. For the Individual Contest, two bonus point problems were predesignated to tutors. Students were not told which problems would earn a bonus point until all students completed their work. At the end of the lesson, tutors tallied checkmarks and awarded students a "half dollar" for each checkmark earned and each correct bonus problem. Fluency condition (Sprint) students also earned a dollar, if the group matched or exceeded the previous week's score twice during the week, or a half dollar, if they did so once during the week. Conceptual condition (Conditioning) students also earned a dollar if the group reached the end of their game board during the week. On the last session of each week, tutors opened the "Fraction Store," where students spent earnings on small prizes priced at \$7, \$13, or \$20. Students exchanged half dollars for whole dollars to determine which prizes they could afford or saved for more expensive prizes in the future. In Lesson 19, we introduced quarter dollars by increasing the number of bonus problems for the Individual Contest.

Fidelity of Implementing the Intervention

Every intervention session was audiotaped. We randomly sampled 20% of recordings (412 recordings) such that tutor, student, and lesson were sampled comparably. A research assistant listened to each sampled tape while completing a checklist to identify the essential points the tutor implemented. The mean percentage of points addressed was 97.32 (SD = 1.44) in the fluency condition and 97.03 (SD = 1.88) in the conceptual condition, $t(16) = .71$, $p = .486$ (the tutor was the unit of analysis, such that each tutor's mean fidelity in the two conditions was compared). For the 5-min practice component, the mean percentage of points addressed was 100.00 (SD = 0.00) for the fluency condition and 100.00 (SD = 0.00) for the conceptual condition. Two research assistants independently listened to 20% (n = 94) of the 412 recordings to assess concordance. The mean difference in score was 2.91% (*SD* = 3.68).

Procedure

As per the study design and the institutional review board, we did not administer individual assessments or collect demographic data on low-risk classmates. The study occurred in five steps. In August and September, for screening, testers administered the WRAT in large groups and then administered the WASI individually to students who met the WRAT criterion for AR status. In September and October, to assess pretreatment comparability among study groups on fraction knowledge, testers administered NAEP and Fraction Addition and Subtraction in three large-group sessions and administered Fraction Number Line in an individual session. Then the pretest moderator measures were administered in two individual sessions (Fraction Number Line in the first session). SWAN teacher ratings were collected approximately 6 weeks into intervention. Intervention occurred for 12 weeks, 3 times per week for 30 min per session from late October to late February. In early March (within 2 weeks of the end of intervention), we assessed intervention effects by readministering NAEP and Fraction Addition and Subtraction in three large-group sessions and readministering Fraction Number Line in one individual session. All individual testing sessions were audiotaped; 20% of tapes were randomly selected, stratifying by tester, for accuracy checks by an independent scorer. Agreement on test administration and scoring exceeded 98%. Testers were blind to conditions when administering and scoring tests.

Results

See Table 1 for pretest and posttest raw scores on the fraction outcome measures for the AR conditions and for low-risk classmates. In the last three columns, Table 1 also shows the magnitude of the pretest and posttest achievement gaps (with respect to low-risk classmates) for each AR condition. These gaps are expressed as effect sizes (raw score difference in means, divided by the pooled *SD*). Table 2 shows results of models assessing the main effects of intervention. Table 3 shows descriptive data and correlations among WM, processing speed, attentive behavior, language, and number line scores for AR students.

Preliminary Analyses

We conducted preliminary analyses as follows. First, we examined whether risk severity, in terms of whether the dichotomous variable (<15th vs. 16th–34th percentile) interacted with intervention condition and in terms of whether the continuous variable on which risk severity was based (pretest calculations) interacted with intervention condition. There were no such interactions, indicating that the pattern of effects was similar regardless of students' pretest calculation skill. Therefore, we did not include risk severity in our analyses. Relatedly, the homogeneity of regression assumption was met for each outcome measure (i.e., the relation between the intervention and the pretest score, the covariate for that outcome variable, was comparable as a function of AR condition): for number line, $F(2, 237) = 0.53$, $p = .588$; for NAEP, $F(2, 237) = 0.07$, $p = .957$; and for fraction procedures, $F(2, 237) = 0.143$, $p = .866$.

Second, we assessed fraction pretest comparability as a function of AR condition. On all measures, pretest performance of the three AR groups was comparable: for number line, $F(2, 240) = 1.75$, $p = .176$; for NAEP, $F(2, 240) = 0.41$, $p = .666$; and for fraction procedures, $F(2, 240) = 0.46$, $p = .629$. Effect sizes for the three measures, respectively, were as follows: for fluency versus control, 0.13 favoring fluency, 0.06 favoring control, and 0.03 favoring control; for conceptual versus control, 0.16 favoring control, 0.08 favoring conceptual, and 0.12 favoring conceptual; and for fluency versus conceptual, 0.31 favoring fluency, 0.14 favoring conceptual, and .14 favoring conceptual.

Do These Interventions Enhance Fraction Knowledge?

To test our first hypothesis, that both interventions would result in superior learning compared to the business-as-usual condition,

Table 1
Pretest and Posttest Score for At-Risk Students and Pre/Post Achievement Gap With Low-Risk Classmates

Variable	At-risk condition: Fluency (n = 84) *M*	*SD*	Conceptual (n = 79) *M*	*SD*	Control (n = 80) *M*	*SD*	Low-risk (n = 265) *M*	*SD*	Achievement gap (effect size): Low-risk vs. fluency	Low-risk vs. conceptual	Low-risk vs. control
Fraction Number Line											
Pretest	0.28	0.05	0.30	0.07	0.29	0.07	NA				
Posttest	0.19	0.08	0.20	0.07	0.27	0.06					
NAEP–Total											
Pretest	8.38	3.00	8.82	3.23	8.58	3.28	12.60	4.18	−1.08	−0.95	−1.01
Posttest	14.25	3.91	14.64	3.64	12.07	3.56	16.48	4.04	−0.58	−0.47	−1.12
Fraction Calculations											
Pretest	3.24	4.04	3.82	4.42	3.34	3.89	6.74	5.16	−0.72	−0.59	−0.70
Posttest	17.43	7.83	17.84	7.90	8.16	4.71	14.43	7.78	0.39	0.44	−0.89

Note. Fraction Number Line is Hamlett et al. (2011), modeled after Siegler et al. (2011), for 0–2 number lines (low scores indicate stronger performance). NAEP–Total is released items on the National Assessment of Educational Progress (19 easy, medium, and hard fourth-grade and easy eighth-grade released fraction items). Fraction Calculations is Fraction Addition and Fraction Subtraction from the 2011 Fraction Battery (Schumacher et al., 2011). NA = not available.

Table 2
Results From Cross-Classified Partially Nested Models Testing Effects of Intervention Versus Control

Measure	Fraction Number Line				NAEP				Calculations			
	Estimate	*SE*	*t*(*df*)	*p*	Estimate	*SE*	*t*(*df*)	*p*	Estimate	*SE*	*t*(*df*)	*p*
Fixed effects												
Adjusted[a] mean, control	−0.60	0.09	−6.66 (73.6)	<.001	−0.43	0.09	−4.62 (77.1)	<.001	−0.76	0.07	−11.27 (56.7)	<.001
Adjusted[a] mean, fluency	0.39	0.12	3.16 (30)	.004	0.17	0.10	1.66 (82.9)	.101	0.36	0.12	2.91 (29.8)	.007
Adjusted[a] mean, conceptual	0.20	0.11	1.75 (68.6)	.084	0.20	0.10	1.98 (24.8)	.059	0.37	0.11	3.40 (78.3)	.001
Adjusted mean difference												
Fluency vs. control	0.99	0.14	7.12 (44.1)	<.001	0.60	0.13	4.69 (139)	<.001	1.12	0.13	8.42 (37.4)	<.001
Conceptual vs. control	0.80	0.13	6.10 (132)	<.001	0.63	0.13	4.97 (45)	<.001	1.13	0.12	9.40 (117)	<.001
Pretest effect	0.14	0.06	2.48 (182)	.014	0.46	0.05	8.40 (226)	<.001	0.16	0.05	3.30 (197)	.001
Variance components												
Classroom-level variance	0.08	0.05	1.54	.062	0.06	0.04	1.60	.055	0.04	0.03	1.35	.089
Small-group level variance												
Fluency condition[b]	0.11	0.14	0.79	.216	0.02	0.10	0.22	.414	0.21	0.14	1.58	.057
Person-level residual variance												
Control condition	0.50	0.10	5.12	<.001	0.55	0.10	5.66	<.001	0.27	0.05	5.29	<.001
Fluency condition	0.77	0.16	4.84	<.001	0.76	0.12	6.13	<.001	0.60	0.12	4.93	<.001
Conceptual condition	0.82	0.14	5.96	<.001	0.61	0.13	4.64	<.001	0.85	0.14	6.01	<.001

Note. Fraction Number Line is Hamlett et al. (2011), modeled after Siegler et al. (2011), for 0–2 number lines. NAEP is released items on the National Assessment of Educational Progress (19 easy, medium, and hard fourth-grade and easy eighth-grade released fraction items). Calculations is Fraction Addition and Fraction Subtraction from the 2011 Fraction Battery (Schumacher et al., 2011).
[a] $F(3, 239)$ for corrected model; $F(1, 239)$ for intercept; $F(1, 239)$ for pretest; $F(2, 239)$ for condition. [b] 1 = Fluency; 2 = Conceptual; 3 = Control.

we conducted a cross-classified partially nested model for each of the three outcomes. This accounted for nesting at the classroom level for all three conditions and at the small-group tutoring level in the two intervention conditions. That is, students (Level 1 units) were partially nested and cross-classified in small groups (Level 2a unit, occurring only in the treatment arms) and classrooms (Level 2b unit). Partial nesting analyses relied on the procedures of Bauer, Sterba, and Hallfors (2008) and Sterba et al. (2013), in which random effects for nesting at the small-group level are employed only in treatment arm(s) and not in the control arm. This procedure involves estimating a random effect that is toggled into the model for each intervention arm and toggled out of the model for the control arm (by estimating random slopes for dummy treatment predictors at the small-group level but no random intercept at the small-group level). Also, residual (person-level) variance was allowed to differ across all three study arms, which avoids the requirement that the control arm necessarily have a smaller model-implied variance than the treatment arms (Bauer et al., 2008; Sterba et al., 2013). This basic partial nesting multilevel model was expanded to account for the cross-classification by also estimating a random intercept at the classroom level. Intraclass correlations (ICCs), not controlling for pretest, were computed taking the cross-classification into account. Accordingly, classroom ICCs were calculated as

Table 3
*Descriptive Information and Correlations Among Measures for Fluency and Conceptual Practice Students Included in Moderation Analyses (*n *= 163)*

Variable	Raw scores		Standard scores		Correlations				
	M	*SD*	*M*	*SD*	LR	A	PS	L	NL Pre
Working memory									
Listening recall (LR)	9.72	2.96	88.24	15.84					
Attentive (A)	35.37	11.96	NA	NA	.12				
Processing speed (PS)	14.53	3.19	90.28	13.18	.25	.25			
Language (L)	30.39	6.49	44.75	9.02	.24	.44	.26		
Fraction Number Line (NL)									
Pre	0.29	0.06	NA	NA	−.01	−.02	.04	−.16	
Post	0.19	0.09	NA	NA	−.04	−.15	−.23	−.23	.13

Note. Working memory is the Working Memory Test Battery for Children–Listening Recall (Pickering & Gathercole, 2001). Attentive behavior is the Strength and Weaknesses of ADHD Symptoms and Normal Behavior scale (Swanson et al., 2004). Processing speed is Woodcock-Johnson III–Cross Out (Woodcock et al., 2001). Language is Wechsler Abbreviated Scales of Intelligence–Vocabulary (Wechsler, 1999; note that standard scores are T scores with $M = 50$ and $SD = 10$). Fraction Number Line is Hamlett et al. (2011), modeled after Siegler et al. (2011), for 0–2 number lines (lower scores indicate stronger performance). Pre = pretest; Post = posttest; NA = not available.

$$\mathrm{ICC}_c^{(a)} = \frac{\hat{\tau}_{00}^{c}}{\hat{\tau}_{00}^{c} + \hat{\tau}_{11}^{g(a)} + \hat{\sigma}^{2(a)}},$$

where $\hat{\tau}_{00}^{c}$ is the random intercept variance at the classroom level, $\hat{\tau}_{11}^{g(a)}$ is the random effect variance for a particular treatment arm (a) at the small group level, and $\hat{\sigma}^{2(a)}$ is the person-level residual variance in arm a. In the control arm, $\hat{\tau}_{11}^{g(a)}$ is 0. Small-group ICCs in the treatment arms were calculated as

$$\mathrm{ICC}_g^{(a)} = \frac{\hat{\tau}_{11}^{g(a)}}{\hat{\tau}_{00}^{c} + \hat{\tau}_{11}^{g(a)} + \hat{\sigma}^{2(a)}}.$$

Again, in the control arm, $\hat{\tau}_{11}^{g(a)}$ is 0. Additionally, although random effects at the small-group level were estimated for both treatment arms in each model, one of these random effects in each model was estimated at its boundary of 0 (causing an estimation problem); thus, it was fixed to 0.

For number line, NAEP, and calculations, respectively, classroom ICCs (controlling for small-group nesting) were .07, .05, and .04. In the fluency intervention arm, small-group ICCs (controlling for classroom nesting) were .15, 0, and .22 and in the conceptual treatment arm, small-group ICCs (controlling for classroom nesting) were 0, .15, and 0.[4]

Cross-classified partially nested multilevel models were run in SAS Proc Mixed using restricted maximum likelihood estimation (REML); standard errors were corrected for small cluster size bias as in Kenward and Rogers (1997). Degrees of freedom for *t* tests of fixed effects were approximated using the procedure of Kenward and Rogers (*ddfm* = *kr* option in SAS) because they do not have a known reference distribution for complex variance component structures such as fitted here (Bauer et al., 2008).

In this way, we conducted cross-classified partially nested multilevel models for each of the three outcomes (number line, calculations, and NAEP) while controlling for the pretest score on the relevant measure. As shown in Table 2, the effect of each intervention condition versus control was significant ($p < .001$). This indicates that outcome scores, controlled for pretest scores, were stronger for each of the two intervention conditions compared to the control group. All intervention effects remained significant following a Bonferroni adjustment of the conventional α of .05 for 2 contrasts ($\alpha = .025$).

Effect sizes for partial nesting designs have previously been calculated (e.g., Compas et al., 2009) as an across-arm mean difference divided by the *SD* within the control arm only. This implies that treatment effects are measured in terms of standard deviation unit change as computed under no manipulation. Effect sizes favoring the fluency condition over the control group ranged from 0.60 to 1.12; effect sizes favoring the conceptual condition over the control group ranged from 0.63 to 1.13. So in support of the first hypothesis, the fraction learning of each intervention condition exceeded that of the control group on all measures. There were no significant differences between the fluency versus conceptual conditions. On number line, NAEP, and calculations, respectively, effect sizes between the two intervention conditions were 0.24, −0.03, and −0.02 (positive values indicate higher performance for the fluency condition).

Does Number Line Estimation Mediate the Effects of the Intervention Focused on the Measurement Interpretation of Fractions?

To assess our second hypothesis, that improvement in the measurement understanding of fractions mediated the effects of each intervention condition over the control group, we followed Preacher and Hayes (2008; Hayes, 2012), using an ordinary least squares path analytical framework. For the indirect (mediation) effects, we used bootstrapping estimation with 5,000 draws to estimate standard errors and 95% confidence intervals (CIs); CIs that do not cover zero are statistically significant. We used standard scores and ran two parallel models: one contrasting fluency practice against the control condition; the other contrasting conceptual practice against the control condition. We focused on the NAEP outcome, instead of procedural fraction skill, because NAEP is a more multifaceted outcome of fraction knowledge and more widely accepted and highly valued. To index the mediator variable (improvement: pretest to posttest gain in the measurement interpretation of fractions), we relied on the number line task because it is a widely used and accepted measure of this construct (e.g., Siegler & Pyke, 2012; Siegler et al., 2011). We multiplied number line scores by −1 so higher standard scores indicate stronger performance.

Figure 1, in the top panel, shows the model contrasting the effects of the fluency condition versus the control group on the NAEP outcome, controlling for NAEP pretest scores. Path coefficients and standard errors (*SE*) are along the arrows. R^2 for this model was .34, $F(3, 160) = 27.22$, $p < .001$. The model partitioned the total effect of 0.59 into direct and indirect effects. The coefficient for the direct effect of 0.47 was significant, $t = 3.50$, $p < .001$ (a 21.7% reduction in the total effect). The coefficient for the indirect effect of 0.12 was also significant, 95% CI [.0442, .2536]. Thus, improvement in measurement understanding partially mediated the intervention effects between the fluency practice and control condition on NAEP. Other coefficients were as follows: constant, 0.31, $t = 3.15$, $p = .002$, and pretest NAEP, 0.43, $t = 6.52$, $p < .001$.

Figure 1, in the bottom panel, shows the model contrasting the effects of the conceptual condition versus the control group on the NAEP outcome, controlling for NAEP pretest scores. R^2 was .38, $F(3, 155) = 31.54$, $p < .001$. The total effect of 0.63 was partitioned into a direct effect of 0.44, $t = 3.23$, $p = .002$ (a 30.2% reduction in the total effect) and an indirect effect of 0.19 was 95% CI [.0777, .3431]. Thus, improvement in measurement understanding also partially mediated the intervention effects contrasting conceptual practice and the control group on NAEP. Other coefficients were as follows: constant, 0.30, $t = 3.27$, $p = .001$, and pretest NAEP, 0.42, $t = 6.93$, $p < .001$.

[4] This pattern of ICCs, found in the cross-classified analysis in SAS, was replicated in a basic partial nesting analysis (ignoring the classroom level of nesting) as well as in the Mplus multilevel module and the SPSS MIXED module.

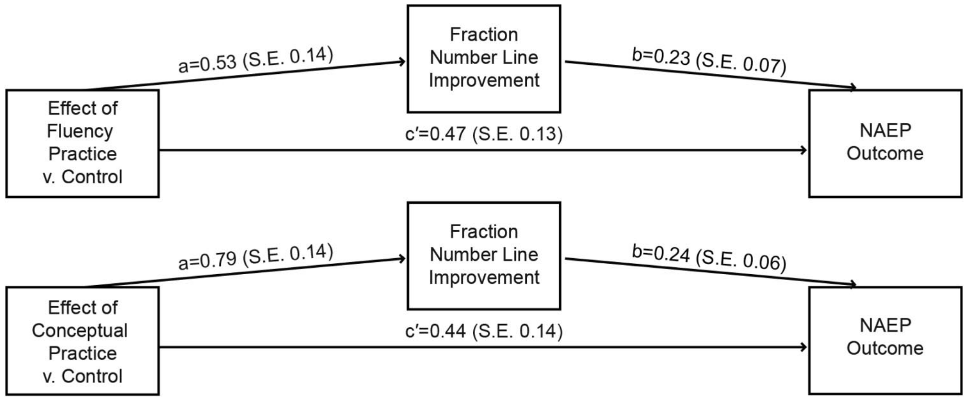

Figure 1. Fraction number line improvement partially mediating the effects of intervention versus control on the NAEP outcome (number line scores multiplied by −1). NAEP = National Assessment of Educational Progress; S.E. = standard error.

Do Individual Differences in WM Moderate Intervention Effects?

To test our third, major hypothesis, concerning the fluency practice compensation effect, we again followed Preacher and Hayes (2008; Hayes, 2012), using an ordinary least squares path analytical framework to estimate whether pretest WM moderated the effects of the fluency practice versus conceptual practice conditions on the number line outcome. We focused on improvement in the number line outcome because, as a mediator of the intervention effect on the NAEP outcome (as in the present study and Fuchs, Schumacher, et al., 2013), it represents an important outcome. It is also a valued outcome in and of itself, given its strong predictive utility (e.g., Siegler et al., 2011). Moreover, because number line estimation accuracy is the most proximal of our outcomes to the activities incorporated in the fluency and conceptual conditions, our hypothesis was that moderation would occur specifically on the number line measure. Again, scores on the number line task were reversed so higher standard scores indicate stronger performance.

In the model we used to test the moderation hypothesis, we controlled for the effects of pretest number line performance, pretest processing speed, pretest attentive behavior, pretest language, and the moderator effects for processing speed, attentive behavior, and language. We controlled for processing speed and attentive behavior, because these constructs are connected to WM. We controlled for language to rule out the possibility that the listening recall WM measure was indexing language ability.[5]

R^2 for the model was .18, $F(10, 152) = 3.29$, $p < .001$, with the following coefficients: for the constant, 0.28 ($SE = 0.07$), $t = 3.94$, $p =< .001$; for pretest WM, −0.01 ($SE = 0.08$), $t = -0.12$, $p = .905$; for the intervention effect (between fluency and conceptual practice), 0.08 ($SE = 0.15$), $t = 0.54$, $p = .590$; for the interaction between WM and intervention, 0.38 ($SE = 0.15$), $t = 2.50$, $p = .013$; for pretest number line estimation accuracy, 0.17 ($SE = 0.08$), $t = 2.21$, $p = .028$; for pretest processing speed, 0.08 ($SE = 0.11$), $t = 0.79$, $p = .433$; for pretest attentive behavior, −0.03 ($SE = 0.12$), $t = -0.25$, $p = .805$; for pretest language, 0.20 ($SE = 0.12$), $t = 1.74$, $p = .085$; for the moderating effect of processing speed, 0.20 ($SE = 0.16$), $t = 1.23$, $p = .220$; for the moderating effect of attentive behavior, 0.12 ($SE = 0.16$), $t = 0.73$, $p = .465$; and for the moderating effect of language, −0.03 ($SE = 0.16$), $t = -0.17$, $p = .864$.

To probe the significant interaction, we relied on the Johnson–Neyman technique (e.g., Bauer & Curran, 2005; Hayes & Matthes, 2009), which derives the value along the full continuum of the moderator at which the effect of *X* on *Y* transitions between statistically significant and not significant. These values demarcate the regions of significance of the intervention effect (between the fluency and conceptual conditions). Thus, this approach does not require arbitrarily operationalizing low, moderate, or high values of the moderator.

In Figure 2, we provide a visualization of the significant interaction. The WM percentile ranks in Figure 2, which represent at-risk within-sample scores, are provided only for the purpose of visualizing the nature of the moderator effect. As explained, the Johnson–Neyman technique derives the value along the full continuum of WM scores at which the intervention effect transitions between significant and not significant, and the shaded areas in the figure indicate the regions of significance. As shown, for students with very low WM, effects favored the conceptual condition; however, for students with more adequate WM, effects favored the fluency condition. The effect favoring conceptual practice transitioned to nonsignificance at 1.93 *SD*s below the AR sample mean on WM. It remained nonsignificant until 0.73 *SD*s above the AR sample mean on WM, at which point the effect transitioned to significance favoring the fluency over conceptual condition. R^2 change due to this interaction was .05 (accounting for 26% of the explained variance), $F(1, 159) = 8.72$, $p = .004$. Thus, WM moderated the effect of intervention between the fluency and conceptual conditions, but the pattern of effect differed from the hypothesized fluency practice compensation effect.

[5] In preliminary analyses, we conducted simple tests (without controlling for competing moderators) for each of the moderating effects. Moderator effects were significant for WM and processing speed, $t = 2.96$, $p = .004$, and $t = 2.10$, $p = .037$, respectively, but not for attentive behavior or language, $t = 1.18$, $p = .240$, and $t = 1.09$, $p = .275$, respectively.

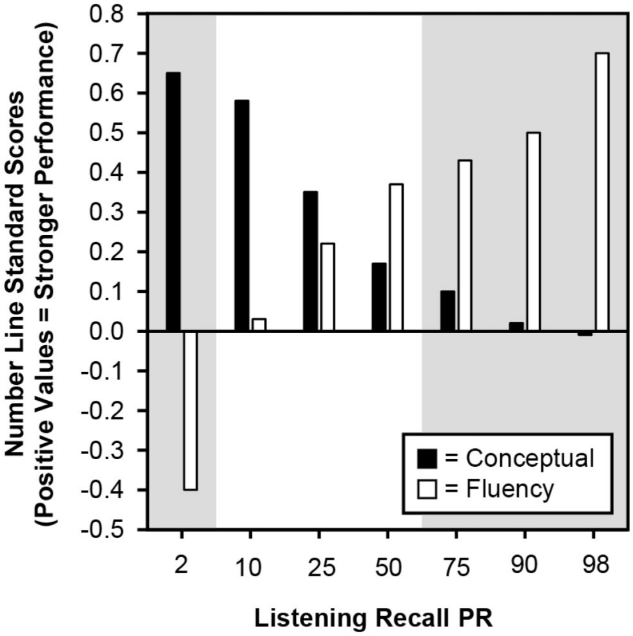

Figure 2. Visualization of working memory (assessed with the Working Memory Test Battery for Children—Listening Recall; Pickering & Gathercole, 2001) as a moderator of the intervention effect between the fluency and conceptual conditions on fraction number line standard scores (positive scores indicate stronger performance). Shaded areas indicate the actual regions of significance. PR = at-risk sample-based percentile rank.

Discussion

Our main purpose in the present study was to assess the fluency practice compensation hypothesis, by examining whether WM moderates the effects of two versions of fraction intervention for AR fourth graders. We began by examining whether each of the two intervention conditions exceeded the control group and whether effects on the highly valued NAEP outcome were mediated by improvement in children's measurement interpretation of fractions. We found that each of the interventions, both focused mainly on the measurement interpretation of fractions, outperformed the control group with large effect sizes, ranging from 0.60 to 1.12 for fluency practice over the control group and from 0.63 to 1.13 for conceptual practice over control. These large effect sizes are especially notable because the control group's core program is a "potentially effective program," according to the What Works Clearinghouse, with stronger outcomes than contrasting core programs (Resendez & Azin, 2008).

Also notable is that the advantage for both intervention conditions included the NAEP, which assessed measurement and part–whole interpretation with comparable emphasis. Moreover, the advantage for both intervention conditions also included fraction calculations, even though the control group focused somewhat more on this topic than the intervention conditions. This supports previous findings suggesting that understanding of fractions is important for learning accurate fraction procedures (e.g., Hecht et al., 2003; Mazzocco & Devlin, 2008; Ni & Zhou; 2005; Rittle-Johnson et al., 2001). Also noteworthy was that, for both intervention conditions, the achievement gap with respect to low-risk classmates was reduced by approximately 50% on NAEP and completely closed for calculations, while the gap for control students widened on both measures. In these ways, each intervention condition resulted in superior fraction learning for AR fourth graders compared to the business-as-usual at-risk control group.

In addition, improvement in the measurement understanding of fractions, as emphasized in both intervention conditions, partially mediated the effect on the NAEP items. This was true for the effect between the fluency condition versus the control group and for the conceptual condition versus the control group. The measurement interpretation of fractions is less intuitive than the part–whole interpretation, which has dominated American schooling to date and was thus emphasized in the business-as-usual control group. The measurement interpretation, by contrast, reflects cardinal size (Hecht, 1998; Hecht et al., 2003), is often (but not exclusively) represented with number lines (e.g., Siegler et al., 2011), and depends on formal instruction. The NMAP (2008) hypothesized that improvement in measurement interpretation is a key mechanism explaining the development of fraction knowledge and recommended that fraction instruction be reoriented in this direction. Finding that this interpretation mediated fraction knowledge at fourth grade on NAEP items, which assess both measurement and part–whole understanding, supports NMAP's hypothesis. It also corroborates Fuchs, Schumacher, et al. (2013), who more specifically showed that improvement in the measurement interpretation mediated effects of intervention versus control group instruction, but students' improvement in part–whole understanding did not. We do, however, remind readers that mediation analyses are correlational; therefore, causation should be attributed with caution.

In terms of the main purpose of the present study, findings revealed an aptitude–treatment interaction, in which WM moderated the effects of the fluency versus conceptual condition on the number line measure. Supplementary conceptual work was superior to the fluency condition for students with very weak WM, but supplementary fluency work promoted better learning for students with more adequate WM. For example, at the AR sample's 10th percentile on WM, the effect size favoring the conceptual condition was strong (0.61). The opposite was true at the AR sample's 90th percentile on WM, with an effect size of 0.52 favoring the fluency condition. Yet, as shown in the lower shaded region of significance in Figure 2, the conceptual condition was advantageous for only a small proportion of the AR sample, which represents an even smaller proportion of a representative sample (because the AR sample, as a whole, was low on WM, scoring a standard score of 88.24 relative to a representative sample; see Table 3).

It is, however, important to note that our model for evaluating WM's moderation effect was stringent, by controlling for processing speed and attentive behavior as well as their moderating effects, even though processing speed and attentive behavior are constructs related to WM. When we ran models that excluded processing speed and attentive behavior and their moderating effects (but included language and its moderating effect), the conceptual condition was significantly more effective beginning at a less extreme point along the distribution of WM scores: 1.57 *SD*s WM (instead of 2.00 *SD*s) below the AR sample mean.

Nevertheless, caution in interpreting this effect is warranted because results did not corroborate the fluency practice compensation hypothesis, as indicated in Fuchs, Geary, et al. (2013). Rather, fluency practice appeared compensatory only for students with more adequate but still low WM, relative to a normative sample. Even so, a much larger proportion of the AR sample benefitted differentially from fluency than conceptual practice. As shown in the upper shaded region of significance in Figure 2, the effect transitioned to significance in favor of fluency practice at 0.73 *SD*s above the AR sample mean, and this includes a larger proportion of a representative sample. A study with a parallel design, with participants who represent the full WM distribution, is needed to examine whether the effect favoring fluency practice transitions back to nonsignificance for students with stronger WM. Such a finding would provide support for the fluency practice compensation hypothesis, but with a complicated pattern in which students with very low WM benefit from supplementary conceptual work, students with more adequate but somewhat low WM benefit from supplementary fluency work, and those with strong WM profit comparably from both forms of supplementary work.

We also remind readers that such moderation analyses are correlational, so as with the mediation analyses, inferences about causation should be applied with caution. Also, although this study's dramatic illustration of an aptitude–treatment interaction suggests the need to personalize intervention, corroborating studies are required. Moreover, research is needed to address the challenges of personalizing instruction, when many AR students will experience limitations across multiple cognitive areas.

In any case, it is instructive that, as reflected in the main effects analyses, number line performance was comparable for the fluency and conceptual conditions. This would indicate that the forms of supplementary work are comparably effective, when incorporated within a larger intervention designed to enhance understanding of fractions generally and measurement interpretation specifically and to build strategic procedural competence on topics related to such understanding. If guided instead by the effect size of 0.24 favoring fluency practice on number line performance (averaged over the full sample), the conclusion would be that fluency practice is superior to conceptual practice. By contrast, the analysis that included the interaction between WM and intervention produces a dramatically different conclusion: Five minutes of supplementary work should focus on consolidating understanding for AR children with very weak WM but should focus on development fluency for AR children with more adequate WM.

This finding has potentially important implications for practice, and it is interesting to consider why different forms of supplementary work produce varying effects, depending on children's WM capacity. In the present study, as part of the larger intervention, children in both conditions received instruction focused dominantly on fraction understanding; the larger program also taught procedural strategies for executing measurement interpretation tasks. These strategies, for example, taught children to segment fraction comparisons into a series of steps, each of which is less resource demanding, and to consider fractional values in relation to benchmark fractions, such as one half. But only students in the fluency condition completed speeded activities designed to build automaticity with such strategies (e.g., developing fluency retrieving from memory fraction equivalents to 1/2). The goal was to reduce demands on (or compensate for poor) WM.

Results, however, suggest that these chunking and benchmarking strategies, which we first taught to all intervention students and then strived to produce automaticity with only in the fluency condition, in fact require some level of WM capacity. After all, consider the fraction comparison chunking/benchmarking strategies, as applied to comparing 5/12 and 3/4, and assume that fluency practice students automatically retrieve fraction equivalents to 1/2 from memory. First, students consider 5/12 by deciding 5/12 should be compared to the closest fraction equivalent to 1/2, identifying the denominator of that fraction equivalent, and retrieving 6/12. Then they compare 5/12 to 6/12 to determine 5/12 is less than 1/2. They store this piece of information in WM while next processing the next fraction in a similar way: deciding 3/4 should be compared to the closest fraction equivalent to 1/2; identifying the denominator of that fraction equivalent, and retrieving 2/4; and comparing 3/4 to 2/4 to determine 3/4 is greater than 2/4. They store this piece of information. Finally, students compare the two pieces of information stored in WM: "3/4 is greater than 1/2" versus "5/12 is less than 1/2" to determine that 3/4 is larger than 5/12.

Some students in the fluency condition did in fact perform better on the (untimed) number line task than did conceptual practice students who received the same intervention without speeded practice on the strategies involved in such WM-demanding fraction comparison tasks. This advantage applied only to fluency condition students with more adequate WM. Thus, for students with the necessary cognitive capacity, in this case WM, the supplementary fluency condition had demonstrable added value. But for students without that necessary cognitive capacity, supplementary work designed to consolidate understanding proved more effective. This conceptual condition depended less on the strategies taught in the larger intervention and instead built on another component of the larger intervention: helping students develop concrete representations of fractions and instantiate meaningful explanations for why fraction values differ.[6]

Such aptitude–treatment interactions have long been discussed in the psychology literature (e.g., Snow, 1991) as a means of differentiating treatment based on individual differences. Historically, however, few studies have uncovered such interactions, perhaps due to constraints imposed by the measurement and statistical methods available at those times. Recently, more aptitude–treatment interactions have been reported. For example, working with second and third graders on mathematics equivalence tasks, Fyfe, Rittle-Johnson, and DeCaro (2012) showed that guidance, provided within exploratory activities that precede explicit instruction, is effective for students with low prior domain knowledge but not for those with higher prior domain knowledge. In a similar way, D. Fuchs, Compton, and Fuchs (2012) showed that at-risk first graders' reading comprehension improved more with intervention targeting word-level skill plus reading comprehension than with intervention exclusively targeting word-level skill. This effect, however, applied only to the two thirds of children with the lowest pretest reading scores. Hornickel, Zecker, Bradlow, and Kraus (2012) showed that the phonological awareness and reading performance of children with reading impairment profit from a year of wearing assistive listening devices, with corresponding

[6] Throughout intervention, including supplementary conceptual activities, we relied on McNeil and Jarvin's (2007) instructional principles when designing manipulative activities.

brain changes to the processing of sounds. But this effect occurred only for students with the greatest initial response variability when processing sounds.

The present study adds to this growing literature on aptitude–treatment interactions by focusing on a general cognitive ability, WM, as the individual difference and by focusing on fraction learning as the outcome. The present study also focused on the intervention contrast between supplementary fluency and conceptual activities, both delivered in the context of a more comprehensive intervention focused dominantly on understanding fractions. In this way, the present study adds to the literature not only on aptitude–treatment interactions but also on the fluency practice compensation hypothesis. The closest prior study we identified contrasted the effects of number knowledge tutoring with fluency versus conceptual practice. In that study, Fuchs, Geary, et al. (2013) considered the role of children's reasoning ability as a moderator of the intervention effect on at-risk first graders' arithmetic skill. Children in the conceptual condition who began first grade with low reasoning ability had poor outcomes relative to medium and high reasoning ability peers in the same condition. By contrast, the effects of fluency practice depended less on children's reasoning ability. So in line with fluency practice compensation hypothesis, fluency practice helped at-risk students compensate for weak reasoning ability.

As discussed, results of the present study are more complicated. The aptitude–treatment interaction (Snow, 1991) showed that one condition favors AR students at the lower end of observed values on the cognitive variable, whereas the other condition favors AR students with more adequate cognitive capacity. The most salient difference between the two studies is the developmental outcome of interest. The first-grade study (Fuchs, Geary, et al., 2013) focused on arithmetic skill (i.e., adding and subtracting problems in the 0–12 range). This form of competence, although highly dependent on children's number knowledge (e.g., Baroody, 1999; Koontz & Berch, 1996), may be more amenable to automatization. This may help account for fluency practice's advantage regardless of children's reasoning ability. In contrast, fraction knowledge is more complex, more abstract, and harder to learn. It may therefore be more demanding of the cognitive moderator, WM.

Interestingly, the three recent aptitude–treatment interaction studies discussed above (D. Fuchs et al., 2012; Fyfe et al., 2012; Hornickel et al., 2012) documented interactions involving individual differences in children's prior domain knowledge. Readers may, therefore, wonder if individual differences in prior mathematics performance may be more relevant than WM in moderating intervention effects. Yet, as reported in our preliminary analyses, pretest calculation skill, which has been identified as a predictor of fraction learning (e.g., Hecht et al., 2003; Jordan et al., 2013; Seethaler et al., 2011), did not interact with the present intervention conditions. Inconsistency between the present study and the three earlier investigations may be due to the fact that whole-number and fraction principles differ in fundamental ways. Thus, at the start of fourth grade, the type of prior mathematics knowledge children have achieved is more distal to the fraction learning challenge than is the relation between prior knowledge and the learning challenge in those prior studies. We also caution readers that findings may be generalized only to intervention that contextualizes brief fluency or conceptual activities within a broader fraction intervention that focuses dominantly on the measurement interpretation of fractions and provides students with strategies for chunking measurement interpretation tasks and for relying on benchmark fractions. In addition, as mentioned, present findings require corroboration, specifically for fraction intervention and more generally for other domains of learning.

Even so, present results—when combined with accumulating evidence that the effects of intervention depend on individual differences in cognitive abilities or prior domain knowledge—indicate the importance of considering aptitude–treatment interactions in future studies. Intervention design for AR students and those with reading or mathematics learning disabilities has improved substantially over past decades. Yet, the dominant approach to intervention, in which all children receive the same standard program, fails to meet the needs of 10% of the general population and 25–40% of students with learning disabilities (e.g., O'Connor & Fuchs, 2013). Aptitude–treatment interactions provide a paradigm that may eventually provide the basis for personalizing intervention in ways that expand efficacy.

Finally, although the aptitude–treatment interaction we documented in the present study is interesting to consider and may have implications for the design of intervention to reach a fuller range of learners, findings of the present study also reveal the strength of the broader intervention designed to foster the measurement interpretation of fractions. On number line estimation, NAEP, and fraction calculations, the effects of such intervention were strong, with the achievement gap for AR learners substantially narrowed or eliminated. Moreover, the failure of the typical school mathematics program to address the needs of a substantial majority of AR learners in a more successful manner raises questions about the quality and nature of business-as-usual fraction instruction. This in part explains widespread difficulty with fractions (e.g., NCTM, 2007; Ni, 2001) and highlights the pressing need to improve the quality of fraction instruction and learning in the United States (NMAP, 2008).

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorder* (4th ed.). Washington, DC: Author.

Baddeley, A. D., & Logie, R. H. (1999). The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9781139174909.005

Baroody, A. J. (1999). Children's relational understanding of addition and subtraction. *Cognition and Instruction, 17,* 137–175. doi:10.1207/S1532690XCI170201

Bauer, D. J., & Curran, P. J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research, 40,* 373–400. doi:10.1207/s15327906mbr4003_5

Bauer, D. J., Sterba, S. K., & Hallfors, D. D. (2008). Evaluating group-based interventions when control participants are ungrouped. *Multivariate Behavioral Research, 43,* 210–236. doi:10.1080/00273170802034810

Behr, M. J., Harel, G., Post, T. R., & Lesh, R. (1992). Rational number, ratio, and proportion. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 296–333). New York, NY: Macmillan.

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

Compas, B. E., Forehand, R. E., Keller, G., Champion, J. E., Rakow, A., Reeslund, K. L., . . . Cole, D. A. (2009). Randomized controlled trial of a family cognitive-behavioral preventive intervention for children of depressed parents. *Journal of Consulting and Clinical Psychology, 77,* 1007–1020. doi:10.1037/a0016930

Compton, D. L., Fuchs, L. S., Fuchs, D., Lambert, W., & Hamlett, C. L. (2012). The cognitive and academic profiles of reading and mathematics learning disabilities. *Journal of Learning Disabilities, 45,* 79–95. doi:10.1177/0022219410393012

Fuchs, D., Compton, D. C., & Fuchs, L. S. (2012). *Responsiveness to reading comprehension intervention as a function of the severity of at-risk first graders' incoming reading performance*. Paper presented at the annual meeting of the Pacific Coast Research Conference, LaJolla, CA.

Fuchs, L. S., Compton, D. L., Fuchs, D., Powell, S. R., Schumacher, R. F., Hamlett, C. L., . . . Vukovic, R. K. (2012). Contributions of domain-general cognitive resources and different forms of arithmetic development to pre-algebraic knowledge. *Developmental Psychology, 48,* 1315–1326. doi:10.1037/a0027475

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., & Bryant, J. D. (2010). The contributions of numerosity and domain-general abilities to school readiness. *Child Development, 81,* 1520–1533. doi:10.1111/j.1467-8624.2010.01489.x

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., Seethaler, P. M., . . . Schatschneider, C. (2010). Do different types of school mathematics development depend on different constellations of numerical versus 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., . . . Changas, P. (2013). First-grade number knowledge tutoring with speeded versus non-speeded practice. *Journal of Educational Psychology, 105,* 58–77. doi:10.1037/a0030127

Fuchs, L. S., & Schumacher, R. F. (2011). *Fraction face-off!* L. S. Fuchs, 228 Peabody, Vanderbilt University, Nashville, TN 37203.

Fuchs, L. S., Schumacher, R. F., Long, J., Namkung, J., Hamlett, C. L., Cirino, P. T., . . . Changas, P. (2013). Improving at-risk learners' understanding of fractions. *Journal of Educational Psychology, 105,* 683–700. doi:10.1037/a0032446

Fyfe, E. R., Rittle-Johnson, B., & DeCaro, M. (2012). The effects of feedback during exploratory mathematics problem solving: Prior knowledge matters. *Journal of Educational Psychology, 104,* 1094–1108. doi:10.1037/a0028389

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities, 37,* 4–15.

Geary, D. C., Boykin, A. W., Embretson, S., Reyna, V., Siegler, R., Berch, D. B., & Graban, J. (2008). *Report of the Task Group on Learning Processes*. Retrieved from http://www.ed.gov/about/bdscomm/list/mathpanel/report/learning-processes.pdf

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

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

Hamlett, C. L., Schumacher, R. F., & Fuchs, L. S. (2011). *2011 number line*. L. S. Fuchs, 228 Peabody, Vanderbilt University, Nashville, TN 37203.

Hayes, A. F. (2012). *PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling*. Manuscript submitted for publication.

Hayes, A. F., & Matthes, J. (2009). Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations. *Behavior Research Methods, 41,* 924–936. doi:10.3758/BRM.41.3.924

Hecht, S. A. (1998). Toward an information processing account of individual differences in fraction skills. *Journal of Educational Psychology, 90,* 545–559.

Hecht, S. A., Close, L., & Santisi, M. (2003). Sources of individual differences in fraction skills. *Journal of Experimental Child Psychology, 86,* 277–302. doi:10.1016/j.jecp.2003.08.003

Hecht, S. A., & Vagi, K. J. (2010). Sources of group and individual differences in emerging fraction skills. *Journal of Educational Psychology, 102,* 843–859. doi:10.1037/a0019824

Hornickel, J., Zecker, S. G., Bradlow, A. R., & Kraus, N. (2012). Assistive listening devices drive neuroplasticity in children with dyslexia. *Proceedings of the National Academies of Science, USA, 109,* 16731–16736. doi:10.1073/pnas.1206628109

Jordan, N. C., Hansen, N., Fuchs, L. S., Siegler, R. S., Gersten, R., & Micklos, D. (2013). Developmental predictors of fraction concepts and procedures. *Journal of Experimental Child Psychology, 116,* 45–58. doi:10.1016/j.jecp.2013.02.001

Kenward, M. G., & Rogers, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53,* 983–997. doi:10.2307/2533558

Koontz, K. L., & Berch, D. B. (1996). Identifying simple numerical stimuli: Processing inefficiencies exhibited by arithmetic learning disabled children. *Mathematical Cognition, 2,* 1–23. doi:10.1080/135467996387525

Krajewski, K., & Schneider, W. (2009). Early development of quantity to number–word linkage as a precursor of mathematical school achievement and mathematical difficulties: Findings from a four-year longitudinal study. *Learning and Instruction, 19,* 513–526.

Kroesbergen, E. H., van Luit, J. E. H., Van Lieshout, E. C. D. M., Van Loosbroek, E., & Van de Rijt, B. A. M. (2009). Individual differences in early numeracy: The role of executive functions and subitizing. *Journal of Psychoeducational Assessment, 27,* 226–236.

Mazzocco, M. M. M., & Devlin, K. T. (2008). Parts and holes: Gaps in rational number sense in children with vs. without mathematical learning disability. *Developmental Science, 11,* 681–691.

McNeil, N., & Jarvin, L. (2007). When theories don't add up: Disentangling the manipulatives debate. *Theory into Practice, 46,* 309–316. doi:10.1080/00405840701593899

Mix, K. S., Levine, S. C., & Huttenlocher, J. (1999). Early fraction calculation ability. *Developmental Psychology, 35,* 164–174. doi:10.1037/0012-1649.35.1.164

Montague, M. (2007). Self-regulation and mathematics instruction. *Learning Disabilities Research & Practice, 22,* 75–83. doi:10.1111/j.1540-5826.2007.00232.x

National Council of Teachers of Mathematics. (2007). *Second handbook of research on mathematics teaching and learning*. Washington, DC: Author.

National Mathematics Advisory Panel. (2008). *Foundations for success: Final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.

Ni, Y. (2001). Semantic domains of rational number and acquisition of fraction equivalence. *Contemporary Educational Psychology, 26,* 400–417. doi:10.1006/ceps.2000.1072

Ni, Y., & Zhou, Y. D. (2005). Teaching and learning fraction and rational numbers: The origins and implications of whole number bias. *Educational Psychologist, 40,* 27–52. doi:10.1207/s15326985ep4001_3

O'Connor, R., & Fuchs, L. S. (2013). Responsiveness to intervention in the elementary grades: Implications for early childhood education. In V. Buysse & E. Peisner-Feinberg (Eds.), *Handbook of response to inter-*

vention in early childhood education (pp. 41–56). Baltimore, MD: Brookes.

Passolunghi, M. C., & Pazzaglia, F. (2004). Individual differences in memory updating in relation to arithmetic problem solving. *Learning and Individual Differences, 14,* 219–230. doi:10.1016/j.lindif.2004.03.001

Passolunghi, M. C., & Pazzaglia, F. (2005). A comparison of updating processes in children good or poor in arithmetic word problem-solving. *Learning and Individual Differences, 15,* 257–269. doi:10.1016/j.lindif.2005.03.001

Pearson Education. (2011). *Scott Foresman Addison Wesley enVisionMATH*. San Antonio, TX: Author.

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

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40,* 879–891. doi:10.3758/BRM.40.3.879

Resendez, M., & Azin, M. (2008). *A study on the effects of Pearson's 2009 enVisionMATH program. 2007–2008: First year report*. Jackson, WY: PRES Associates.

Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology, 91,* 175–189. doi:10.1037/0022-0663.91.1.175

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93,* 346–362. doi:10.1037/0022-0663.93.2.346

Schumacher, R. F., Namkung, J. M., Malone, A., & Fuchs, L. S. (2011). *2011 fraction battery*. L. S. Fuchs, 228 Peabody, Vanderbilt University, Nashville, TN 37203.

Seethaler, P. M., Fuchs, L. S., Star, J. R., & Bryant, J. (2011). The cognitive predictors of computational skill with whole versus rational numbers: An exploratory study. *Learning and Individual Differences, 21,* 536–542. doi:10.1016/j.lindif.2011.05.002

Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., . . . Chen, M. (2013). Early predictors of high school mathematics achievement. *Psychological Science, 23,* 691–697. doi:10.1177/0956797612440101

Siegler, R. S., & Pyke, A. A. (2012). Developmental and individual differences in understanding of fractions. *Developmental Psychology*. Advance online publication. doi:10.1037/a0031200

Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology, 62,* 273–296. doi:10.1016/j.cogpsych.2011.03.001

Snow, R. E. (1991). Aptitude–treatment interaction as a framework for research on individual differences in psychotherapy. *Journal of Consulting and Clinical Psychology, 59,* 205–216. doi:10.1037/0022-006X.59.2.205

Sterba, S. K., Preacher, K., Hardcastle, E., Forehand, R., Cole, D., & Compas, B. (2013). *Structural equation modeling approaches for handling partially nested data*. Manuscript in preparation.

Swanson, J., Schuck, S., Mann, M., Carlson, C., Hartman, K., Sergeant, J., . . . McCleary, R. (2004). *Categorical and dimensional definitions and evaluations of symptoms of ADHD: The SNAP and the SWAN rating scales*. Retrieved from http://www.adhd.net/SNAP_SWAN.pdf

Tall, D., Gray, E., Bin Ali, M., Crowley, L., DeMarois, P., McGowen, M., . . . Yusof, Y. (2001). Symbols and the bifurcation between procedural and conceptual thinking. *Canadian Journal of Science, Mathematics and Technology Education, 1,* 81–104. doi:10.1080/14926150109556542

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114,* 104–132. doi:10.1037/0033-295X.114.1.104

U.S. Department of Education, Institute of Education Sciences. (2013, January). *Elementary School Mathematics Intervention report: enVisionMATH*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_envisionmath_011513.pdf

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Psychological Corporation.

Wilkinson, G. S., & Robertson, G. J. (2006). *Wide Range Achievement Test 4 professional manual*. Lutz, FL: Psychological Assessment Resources.

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

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

Received January 7, 2013
Revision received August 2, 2013
Accepted August 7, 2013 ■