Exceptional Children
2019, Vol. 85(2) 147–162

DOI: 10.1177/0014402918782628
journals.sagepub.com/home/ecx
SAGE

Does the Severity of Students' Pre-Intervention Math Deficits Affect Responsiveness to Generally Effective First-Grade Intervention?

Lynn S. Fuchs[1], Douglas Fuchs[1], and Jennifer K. Gilbert[1]

Abstract

The purpose of this analysis was to assess whether effects of first-grade mathematics intervention apply across the range of at-risk learners' initial skill levels. Students were randomly assigned to control (*n* = 213) and two variants of intervention (*n* = 385) designed to improve arithmetic. Of each 30-minute intervention session (48 over 16 weeks), 25 minutes were identical in the two variants, focused on number knowledge that provides the conceptual bases for arithmetic. The other five minutes provided nonspeeded conceptual practice (*n* = 196) or speeded strategic practice (*n* = 199). Contrasts tested effects of intervention (combined across variants) versus control and effects between the variants. Moderation analysis indicated no significant interactions between at-risk children's pre-intervention mathematics skill and either contrast on any outcome. Across pre-intervention math skill, effects favored intervention over control on arithmetic and transfer to double-digit calculations and number knowledge, and favored speeded over nonspeeded practice on arithmetic.

When a randomized control trial (RCT) produces statistically significant effects favoring the learning outcomes of students who receive intervention over those who do not, that intervention is deemed validated. *Validation* suggests most students respond to the intervention, but few, if any, standard (non-individualized) interventions achieve universal response. Some students require adjustments to make intervention more intensive (O'Connor & Fuchs, 2013).

Little is known about student characteristics that explain responsiveness. One possibility is that the robustness of intervention effects depends on the level of students' pre-intervention academic skill. We identified two previous studies that assessed the efficacy of mathematics intervention as a function of students' pre-intervention math performance. In Smith, Cobb, Farran, Cordray, and Munter (2013), first graders who received Math Recovery outperformed the control group on arithmetic, concepts and applications, quantitative concepts, and math reasoning. Effect sizes (ESs) ranged from 0.15 to 0.30, but were *larger* for children who began intervention *below* the 25th percentile (ESs = 0.31–0.40) than for students who exceeded the 25th percentile. This suggests students with greater math competence had less need for Math Recovery. Yet two-thirds of Smith et al.'s "at-risk" sample began intervention scoring above the 25th percentile, and scores ranged as high as the 95th percentile.

By contrast, in a sample selected to represent the distribution of at-risk learners' low mathematics skill (1st–34th percentile, sampled purposely to represent the full distribution), Fuchs,

[1]Vanderbilt University

Corresponding Author:
Lynn S. Fuchs, Vanderbilt University, 228 Peabody, Nashville, TN 37203.
E-mail: lynn.fuchs@vanderbilt.edu

Sterba, Fuchs, and Malone (2016) tested whether pre-intervention math performance moderated the effects of a fourth-grade fractions intervention. No moderation effect was identified: Students benefited comparably from the intervention, with similar magnitude of effects for at-risk intervention students over at-risk control group students across points along pre-intervention whole-number math skill. Although this suggests the robustness of intervention, it may indicate that pre-intervention mathematics performance is not a viable basis for forecasting which students will respond inadequately and are in need of more intensive intervention.

Yet, fractions at fourth grade have some features that may not generalize to other mathematics topics. For example, the principles that guide whole-number understanding (used to index pre-intervention math skill at the start of the Fuchs et al., 2016 intervention) differ from those guiding rational-number thinking (used to index intervention outcome), and the whole-number calculations involved in fourth-grade fractions are relatively simple. Thus, in the initial phases of fractions learning, students do not require strong whole-number knowledge and operational skill to succeed. Clearly, additional research is needed to examine at-risk pre-intervention math skill as a moderator of intervention effects in which the moderator represents a critical foundational skill for the math learning addressed during intervention.

Purpose of the Present Analysis

The purpose of the present analysis was to revisit whether pre-intervention mathematics performance moderates intervention efficacy using a first-grade whole-number intervention designed to improve children's arithmetic (Fuchs et al., 2013). Pre-intervention whole-number understanding and operational skill (the moderator variable) is important for success with arithmetic, and a good distribution of performance on whole-number understanding and skill exists at the start of first grade. We selected the Fuchs et al. (2013) study for moderation analysis for this and three additional reasons.

The second reason was that first-grade arithmetic skill is a critical target. It predicts mathematics learning through the end of fifth grade (Geary, 2011) and eventual mastery of high school algebra, a gateway for later entry into mathematics-intensive fields (U.S. Department of Education, 2008). Third, the screening criteria for entering the study were designed to create a viable distribution of pre-intervention scores at the at-risk end of math performance. Fourth, pre- and postintervention normative data were available on the same pre- and post-intervention measures for not-at-risk classmates. This permitted us not only to look at moderation analysis, in which the performance of intervention students is contrasted to at-risk control group students, but also to consider rates of responsiveness-to-intervention; evaluating at-risk students' growth and end-of-intervention performance against not-at-risk classmates on the same measures.

Results may provide insight into how pre-intervention math performance affects responsiveness to generally effective intervention, with implications for making intervention decisions more efficient and effective. For example, if results show that intervention fails to serve the needs of students with *more severe* pre-intervention math deficits, schools might place these students in intensive intervention immediately, before they experience failure in a standard intervention. Alternatively, if results show that intervention fails to serve the needs of students with *less severe* pre-intervention math deficits, schools might forgo placing these students in intervention and instead focus on adjustments to the general education program.

Information on First-Grade Math Development and Our Approach to Intervention

Before describing study methods and our analytic procedures, we contextualize the Fuchs et al. (2013) study by describing how competence with first-grade arithmetic develops and how our approach to intervention reflects the developmental pathways. When children enter first grade, most have a rudimentary understanding of addition and subtraction and can count to solve problems (Geary, 1994). For addition, young children count both addends; for subtraction, they represent the beginning quantity with objects, sequentially separate the number of objects to be subtracted, and then count the remaining set (Groen & Resnick, 1977). As understanding of cardinality and the counting sequence develop, children discover the number-after rule for adding with 1. They also come to understand that the sum of 5 + 2 is two numbers beyond 5, which leads them to discover the efficiency of counting from the first addend and to rely on more efficient counting procedures. For addition, the most efficient counting procedure involves starting with the cardinal value of the larger addend and counting up the number of times equal to the smaller addend (for 2 + 3 = "two: three, four, five"); for subtraction, starting with the subtrahend and counting to the minuend (for 7 − 4 = "four: five, six, seven"; the answer is the number of counts).

Frequent use of efficient counting procedures reliably produces the correct association between problem and answer, which results in long-term memories (Fuson & Kwon, 1992; Siegler & Robinson, 1982). This enables direct retrieval of answers, and the commutativity of addition facilitates retrieval of related addition problems (Rickard, Healy, & Bourne, 1994). Subtraction, which is not commutative, is more difficult. It is facilitated by retrieval of related addition facts (e.g., 8 − 5 = 3, based on 5 + 3 = 8; LeFevre & Morris, 1999) once children understand the inverse relation between addition and subtraction (Geary, Boykin, Embretson, Reyna, Siegler, Berch, & Graban, 2008).

Difficulty with arithmetic is an indicator of risk for long-term learning disabilities (Geary, Hoard, Nugent, & Bailey, 2012). Students with mathematics learning disabilities show consistent delays in the adoption of efficient counting procedures, make more counting errors during their execution, and fail to make the shift toward memory-based retrieval of answers (e.g., Geary et al., 2012; Goldman, Pellegrino, & Mertz, 1988). Most eventually catch up to peers in skilled use of counting procedures, but difficulty with retrieval tends to persist (Geary et al., 2012; Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; Jordan, Hanich, & Kaplan, 2003).

The major emphasis of the intervention in the present analysis was developing the conceptual bases for arithmetic, as reflected in the developmental pathways described previously. The five minutes of each 30-minute session devoted to practice was nonspeeded (reviewing the conceptual bases underpinning arithmetic problems) or speeded (promoting strategic, quick responding and use of efficient counting procedures to generate many correct responses to arithmetic problems).

In the Fuchs et al. (2013) RCT, we investigated the efficacy of intervention on first graders' competence with arithmetic while assessing transfer to two-digit calculations and an integrated measure of number knowledge. We considered calculations a form of transfer because two-digit calculations was not a major focus of intervention. We considered the number knowledge task transfer because it represented an integrative form of knowledge across multiple dimensions of number knowledge and it was novel, not explicitly taught or practiced during intervention. (We omitted word problems from the present analysis despite significant word-problem effects in Fuchs et al. due to space constraints and because word problems was not a major focus of intervention. They were used only to contextualize number sentences. Also, results paralleled findings for the other transfer tasks.)

For the present analysis, we asked two questions. For the first, "Does the effect of math intervention compared to control differ depending on students' pre-intervention math skill?," we combined the two intervention conditions and estimated the effect between intervention versus control. For the second, "Does the effect between the two types of practice conditions differ depending on pre-intervention mathematics skill?," we compared math intervention with speeded practice to math intervention with nonspeeded practice. For each outcome, the corresponding pre-intervention score was treated as the moderator of treatment effects.

Method

Participants

Additional information on participants and other methods is available in Fuchs et al. (2013). Across four cohorts in four consecutive years, we recruited 40 schools and 233 first-grade classes. We relied on a latent class approach to screen the first cohort of children for at-risk and not-at-risk status by combining scores across math applications, concepts, calculations, and word-reading screeners into a single latent factor, used to designate risk status. For remaining cohorts, we used the first-year cut-points for consistency. We excluded students whose teachers identified them as non-English speakers or with standard scores below 80 on both subtests of the two-subtest Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999).

We enrolled into the study 648 at-risk (below the 40th percentile on the latent factor score) and 325 not-at-risk (above the 40% percentile) students from 227 classes. (The not-at-risk sample was used to set responsiveness criteria. The other six classrooms did not include enough at-risk students to participate.) Then we randomly assigned at-risk students at the individual level, stratifying by pre-intervention math scores and classrooms, to three conditions: control, conceptual arithmetic intervention with speeded practice (A+SP), and conceptual arithmetic intervention with nonspeeded practice (A+NSP).

During first grade, some not-at-risk and at-risk students (distributed comparably across conditions) moved outside the study's reach, leaving 307 not-at-risk and 608 at-risk students in 218 classrooms from 39 schools. Of the 608 at-risk students, 213 were in the control group, 199 in the A+SP condition, and 196 in the A+NSP condition. Sample size for this analysis is slightly larger than reported in Fuchs et al. (2013) (591 vs. 608 at-risk; 300 vs. 307 not-at-risk) because Fuchs et al.'s analyses were restricted to children who were also tested on a battery of cognitive process measures.

See Table 1 for pre-intervention scores on the Wide Range Achievement Test-3-Arithmetic (WRAT; Wilkinson, 1993) and WASI IQ by condition. Control group children were 73.2% African American, 17.4% non-Hispanic White, 6.1% Hispanic White, and 3.3% other. A+SP children were 67.8% African American, 22.6% non-Hispanic White, 6.5% Hispanic White, and 0% Other. A+NSP children were 67.3% African American, 20.4% non-Hispanic White, 8.2% Hispanic White, and 4.1% Other. Not-at-risk children were 39.7% African American, 41.7% non-Hispanic White, 8.5% Hispanic White, and 10.1% Other.

Intervention

Intervention, which addresses the conceptual and procedural bases for emerging competence with arithmetic, occurred three times per week, 30 minutes per session, for 16 weeks in a quiet location outside of classrooms. Make-ups ensured 48 sessions. The program is organized in a manual (*Galaxy Math*; Fuchs, Fuchs, & Bryant, 2010) with materials and guides that provide each lesson's structure, content, and language of explanation. To ensure the natural flow of interactions and responsiveness to student difficulties, tutors review but do not read from or memorize lesson guides. To foster engagement, the program uses a space theme. Each lesson includes a 25-minute segment on the conceptual bases for arithmetic and five minutes of practice to support accurate arithmetic skill. Content and activities in the 25-minute segment were the same in the two practice conditions.

Lessons are organized in five units. Unit 1 addresses basic number knowledge; Unit 2, arithmetic doubles; Unit 3, arithmetic sets 5 through 12 (e.g., the 5 set includes all problems with sums or minuends of 5); and Unit 4, 10s concepts. Students who advance quickly through most lessons also complete Unit 5, a review set of lessons. Instruction incorporates manipulatives and number lines (1–19 through Unit 3; 1–100 for Unit 4).

Unit 3, which comprises approximately half the program, focuses on partitioning numbers into constituent sets and number families (e.g., for the 5 set, 0 + 5, 1 + 4, 2 + 3,

Table 1. Descriptive Pre-Intervention Performance and Pre- and Posttreatment Outcomes for At-Risk Students by Condition and Not-At-Risk Students.

	At Risk							
	Control		A+SP		A+NSP		Not at Risk	
Variable	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Descriptive								
WRAT	88.81	12.05	89.15	11.67	89.68	12.76	107.29	11.95
WASI IQ	85.63	8.07	85.76	7.81	85.46	8.45	100.37	13.34
Outcomes								
Arithmetic-pre	12.42	7.31	12.56	7.58	12.47	6.78	29.91	11.85
Post	22.18	11.58	33.27	14.02	27.97	11.82	42.73	14.21
Calculations-pre	2.66	2.71	2.86	2.85	2.80	2.83	8.64	6.59
Post	6.08	5.89	10.28	7.78	9.18	6.76	15.45	9.72
Number knowledge-pre	−0.51	0.72	−0.50	0.83	−0.49	0.67	1.02	1.02
Post	−0.76	1.23	−0.37	1.34	−0.49	1.22	1.09	1.07

Note: A+NSP = Arithmetic plus nonspeeded practice; A+SP = Arithmetic plus speeded practice; WASI = Wechsler Abbreviated Scale of Intelligence; WRAT = Wide Range Achievement Test-3-Arithmetic.

$5 - 0, 5 - 1, 5 - 2$, etc.), with five activities per lesson. First, tutors and students use unifix cubes to explore how the target number can be partitioned in different ways to derive the addition and subtraction problems in that set. The second activity focuses on number families in that set, with visual displays that group families in the set and blocks to help students rely on part-whole knowledge to understand why four problems make families. Third, for the target set, students generate all addition and subtraction problems, using rockets to show problems. Fourth, tutors and students create a story problem together on that set, produce the answer, and explain why the problem belongs in the set. Fifth, students review previous sets with corrective feedback. Between one and four lessons occur in each set; mastery criteria determine the pace with which students move through sets.

The content addressed in the final 5-minute practice segment of each lesson was the same in both practice conditions: that day's number knowledge lesson (through Lesson 4: number identification, magnitudes, greater or less than; after Lesson 4: sets with ±0, 1, and 2, and then sets with sums and minuends of 5–12). Practice activities differed by condition. Nonspeeded practice reinforces thoughtful application of the relations and principles that serve as reasoning strategies to support arithmetic procedural skill. Students use space theme manipulatives to play games that provide contextualized review of that day's lesson. For example, for $n \pm 1$ lessons, children spin a dial segmented from 1 to 19 to identify the number of "rockets called to explore the math galaxy" and count this number of rockets onto the game board. Then, the tutor informs the child that one more rocket is needed or one is called back to the space station, so the child adds or takes one away. The child then states the number sentence with answer. For an 8 set lesson, children are informed how many rockets constitute the fleet and write that numeral as the total. Then they roll a die to find the first group of rockets released from the space station, count that number onto the game board, and write the numeral as an addend. Then, they determine how many more rockets are needed to complete the fleet, write that numeral as an addend, and read the number sentence. Next, they roll the die to find how many rockets are called back to the space station and write numerals to generate and read a number sentence. Games differ for each day on the same topic. In nonspeeded practice and lessons, tutors encourage children to know the answer or rely on number principle strategies, including using number lines, arithmetic principles, and efficient counting strategies. "Knowing the answer right off the bat" is the preferred strategy when students are sure of answers.

Speeded practice promotes quick responding and use of efficient counting procedures to generate many correct answers, with the goal of forming long-term memory representations for retrieval. Students complete the "Meet or Beat Your Score" game, which provides 90 seconds to answer flash cards (e.g., for $n \pm 1$ lessons, flashcards are $n + 1$ and $1 + n$ problems where $n = 0–18$; for 8 set lessons, flashcards are addition problems with the sum 8 and subtraction problems with the minuend 8). Each presented problem is answered correctly because as soon as an error occurs, tutors require children to use the taught counting strategy to produce the correct response. Time elapses as children use the counting procedure (as many times as needed). So, careful but quick responding increases the number of correct responses, which is charted on a Rocket Chart at the end of 90 seconds. Then, the child has two chances to meet or beat that score. In speeded practice and lessons, tutors require children to know the answer (retrieve) or use the efficient counting strategies they have been taught. "Knowing the answer right off the bat" is preferred if the child is confident of the answer.

Measures and Data Collection

To index arithmetic skill, we used *Arithmetic Combinations* (Fuchs, Hamlett, & Powell, 2003), with Addition (25 problems, sums 5 to 12) and Subtraction (25 problems, minuends 5 to 12). For each subtest, students have one minute to write answers. We used total number of correct answers across Addition and Subtraction. On this sample, α was .96. To index transfer to more complex calculations, we used *Double-Digit Addition and Subtraction* (Fuchs et al., 2003), with two subtests: Addition (20 two-digit addition problems with and without regrouping) and Subtraction (20 two-digit subtraction problems with and without regrouping). For each subtest, students have five minutes to write answers. On this sample, α was .94.

To assess transfer to an integrative number knowledge task, we used the Number Sets Test (Geary, Bailey, & Hoard, 2009), which indexes the speed and accuracy with which children understand and operate with small numerosities (<10) while transcoding between quantities and symbols. At the top of the page, the target sum (5 or 9) is shown. For each target sum, dominoes on the first page contain arrays of objects with same or different objects; the second page shows objects with Arabic numerals and Arabic numerals with Arabic numerals. The child circles groups that combine to make the target sum, with 60 seconds per page for the sum 5 and 90 seconds for the sum 9. Signal detection methods are applied to the number of hits and false alarms to generate a d' variable representing the child's sensitivity to quantities (Geary et al., 2009). Test-retest reliability on a sample of 50 participants was .89.

Fidelity of Implementation and Data Collection

Every intervention session was audiotaped. We randomly sampled 20% of 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. Agreement exceeded 97%. Research assistants, unfamiliar to the children they tested, administered measures in groups. We audiotaped individual test sessions and rescored 20% of recordings. Agreement exceeded 98%.

Data Analysis

For the first question, "Does the effect of number knowledge intervention compared to control differ depending on students' pre-intervention skill?", we compared the effect across the two intervention conditions against control. For the second question, "Does the difference in effect between the two types of number knowledge intervention differ depending on pre-intervention skill?", we compared A+SP against A+NSP. For each outcome, the corresponding pre-intervention score was treated as the moderator of the intervention effect. To answer our questions, we used two orthogonal contrast codes. The estimate of the first, c1_TvC (control = $-.\overline{66}$, A+SP = $.\overline{33}$, and A+NSP= $.\overline{33}$), represents the mean difference between students in intervention versus control. The estimate of the second, c2_A+SPvA+NSP (control = 0, A+SP = .5, and A+NSP = −.5), represents the mean difference between A+SP versus A+NSP.

Prior to running moderation analyses, outcome data were screened for nonnormality and extreme values. One posttreatment arithmetic score, nearly 4 standard deviation (*SD*) above the mean, was winsorized to the next closest value. Four scores on posttreatment number sense were at least 2.5 *SD* below the mean and discrepant from the remainder of the distribution. These scores were winsorized to the next closest values. Then, pre-intervention comparability among conditions was examined with three analysis of variance (ANOVA) models. No significant differences were detected among groups for arithmetic, $F(2, 605) = 0.02$, $p = .982$; (square root of) calculations, $F(2, 605) = 0.88$, $p = .415$; or number knowledge, $F(2, 605) = 0.05$, $p = .947$.

Our data structure incorporated three levels: students (Level 1), cross-classified by classrooms and teachers (Level 2), and classrooms and teachers nested in schools (Level 3). For each outcome, we ran unconditional multilevel models including a random effect for classrooms, teachers, and schools to judge the necessity of including each in the final model. Further, because we assumed that residual variance might vary by condition, we estimated separate residual variances for each. In the final model, we retained all nonzero random effects. We relied on likelihood ratio tests to signal the need for heteroscedastic or homoscedastic residuals by condition.

For the three moderator models, pre-intervention variables were grand-mean-centered before generating interaction variables. Interaction variables were calculated by multiplying the centered pre-intervention score (the same measure as the outcome but measured prior to intervention) by both contrast codes. Equation 1 represents the final generic model:

$$y = \beta_0 + \beta_1 * cPre + \beta_2 * c1\_TvC + \beta_3 * c2\_DPvG + \beta_4 * cPre * c1\_TvC + \beta_5 * cPre * c2\_DPvG + u_{0k} + r_{0(j1)k} + r_{0(j2)k} + e_{i(j1)(j2)k}, \quad (1)$$

where β_0 is the intercept (the average post-intervention score across conditions for students with the average pre-intervention score); β_1 is the effect of the pre-intervention variable; β_2 is the mean difference between intervention and control conditions, controlling for pre-intervention scores; β_3 is the mean difference between A+SP and A+NSP conditions, controlling for pre-intervention scores; β_4 is the interaction effect between pre-intervention scores and Contrast 1, intervention versus control; β_5 is the interaction effect between pre-intervention scores and Contrast 2, A+SP versus A+NSP; u_{0k} is the random residual for school *k*; $r_{0(j1)k}$ is the random residual for classroom *j*(1) in school *k*; $r_{0(j2)k}$ is the random residual for teacher *j*(2) in school *k*; and $e_{i(j1)(j2)k}$ is the random residual for student *i* in classroom *j*(1) taught by teacher *j*(2) in school *k*. The *j* subscripts are assigned to both classrooms and teachers to signify that those effects are crossed at the same level. This model assumes homoscedasticity of Level 1 residuals across conditions. However, separate variance components across conditions were estimated if determined necessary by likelihood ratio tests (Sterba, 2017).

Moderator models were first run using Stata's *quietly* command in which results were hidden from view but residuals were available for inspection. Level 1 residuals were examined for violations of normality and homoscedasticity. We ran the final models and obtained results only after making necessary remediations, as described in the Results section.

Results

Multilevel Multivariate Inferential Models

Arithmetic. In the initial arithmetic model, three multivariate outliers were detected. Those cases were removed before estimating a final moderator model, in which the Level 1 residuals met assumptions of normality and homoscedasticity. Neither interaction was statistically significant at the α = .05 level, β_4 = 0.05, *SE* = 0.10, *p* = .597, and β_5 = 0.21, *SE* = 0.14, *p* = .137. In Figure 1, we graphed the nonsignificant interactions. The parallel lines illustrate similar intervention effects across the distribution of pre-intervention scores. The bottom graph shows a slightly larger effect of A+SP over A+NSP at the upper end of the pre-intervention arithmetic distribution than at the lower end, but again this interaction (*p* = .137) was not statistically significant.

Because neither interaction was significant, we removed both and reran a main effects model to calculate treatment ESs (Hedges's *g*s based on model coefficients; U.S. Department of Education, 2013), with significant effects for both contrasts, β_2 = 8.54, *SE* = 0.72, *p* < .001, ES = 0.69 (mean of both intervention conditions vs. control), and β_3 = 5.65, *SE* = 1.02, *p* < .001, ES = 0.44 (A+SP vs. A+NSP). We accounted for the potential inflation of Type I error due to multiple comparisons across contrasts and outcomes by using the Benjamini-Hochberg (Benjamini & Hochberg, 1995) false discovery rate to adjust critical *p* values. The main effects remained significant even with the correction. Table 1 shows simple pre- and post-intervention means and *SD*s by condition on all three outcomes.

Calculations. Residuals from the initial calculations model failed normality and homoscedasticity assumptions. Pre- and post-intervention calculations variables had right-skewed distributions, so the square root function was applied, resulting in more normal distributions. The model was rerun using the transformed variables; Level 1 residuals from this model were normal and homoscedastic. Neither interaction was statistically significant at the α = .05 level, β_4 = −0.04, *SE* = 0.11, *p* = .730, and β_5 = −0.04, *SE* = 0.13, *p* = .767. In Figure 2, parallel lines in both graphs illustrate these nonsignificant interactions.

Interaction terms were removed from the model to run a main effects model, with a significant effect for Contrast 1, β_2 = 0.66, *SE* = 0.09, *p* < .001, ES = 0.53 (both intervention conditions vs. control), but not Contrast 2, β_3 = 0.22, *SE* = 0.11, *p* = .047, ES = 0.17 (A+SP vs. A+NSP), after adjusting critical *p* value cutoffs to control the false discovery rate (Benjamini & Hochberg, 1995). Thus, on the calculations outcome, the difference between intervention and control was reliable but not the difference between the two practice conditions.

Number knowledge. Initial models produced seven multivariate outliers as detected by standardized residuals >|3|. Even after omitting these values, the residuals appeared somewhat heteroscedastic, so we employed the Huber-White sandwich estimator to correct standard errors in the final model. On arithmetic and calculations, neither interaction was statistically significant at the α = .05 level, β_4 = 0.01, *SE* = 0.11, *p* = .920, and β_5 = −0.02, *SE* = 0.17, *p* = .895. Figure 3 illustrates this finding. The main effects model without the interaction terms revealed a significant main effect for Contrast 1, β_2 = 0.33, *SE* = 0.08, *p* < .001, ES = 0.27 (both intervention conditions vs. control), but not Contrast 2, β_3 = 0.14, *SE* = 0.10, *p* = .134, ES = 0.11 (A+SP vs. A+NSP).

Responsiveness to Intervention in Terms of Growth and Post-Intervention Level

We also considered the proportion of children classified as inadequately responsive to intervention. In these analyses, we operationalized inadequate response in two ways: growth and final status, both based on the normalization principle (Frijters, Lovett, Sevcik, & Morris,

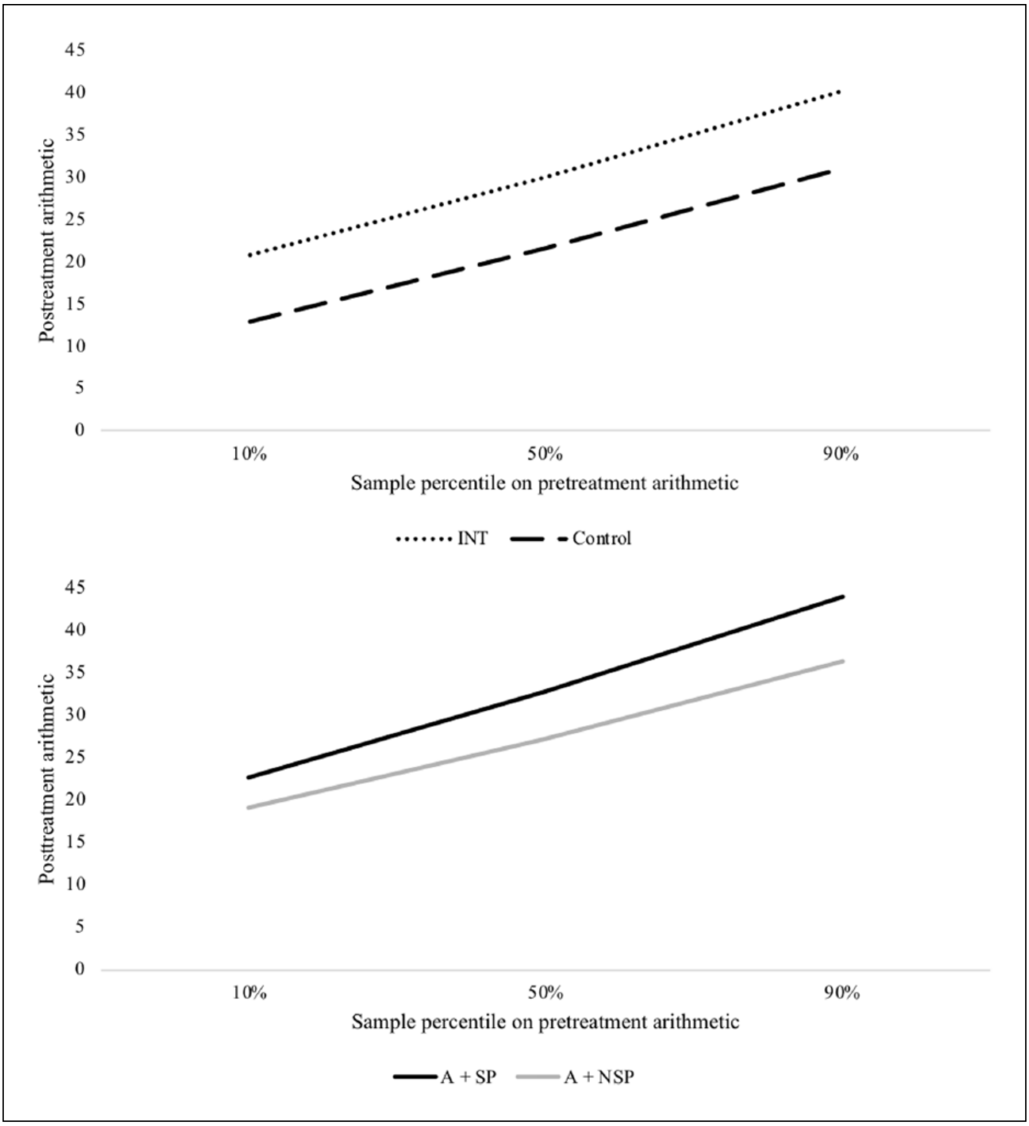

Figure 1. Graphical illustration of nonsignificant interaction effects on arithmetic: Top panel is intervention versus control conditions; bottom panel is A+SP (arithmetic concepts plus speeded practice) versus A+NSP (arithmetic concepts plus nonspeeded) conditions.

2013; Fuchs, 2003). For *growth*, inadequate response was defined as improvement (post-intervention performance minus pre-intervention performance) below the 25th percentile of the not-at-risk classmates' distribution of improvement scores. For *final status*, inadequate response was defined as post-intervention score below the 25th percentile of the not-at-risk classmates' distribution of post-intervention scores.

On arithmetic, inadequate growth occurred for 31% of control students, 6% of A+SP students, and 13% of A+NSP students. By contrast, inadequate post-intervention performance level occurred for 84% of control students, 48% of A+SP students, and 64% of A+NSP students. On the calculations transfer outcome, inadequate growth occurred for 45% of control students, 22% of A+SP students, and 26% of A+NSP students. By contrast, inadequate

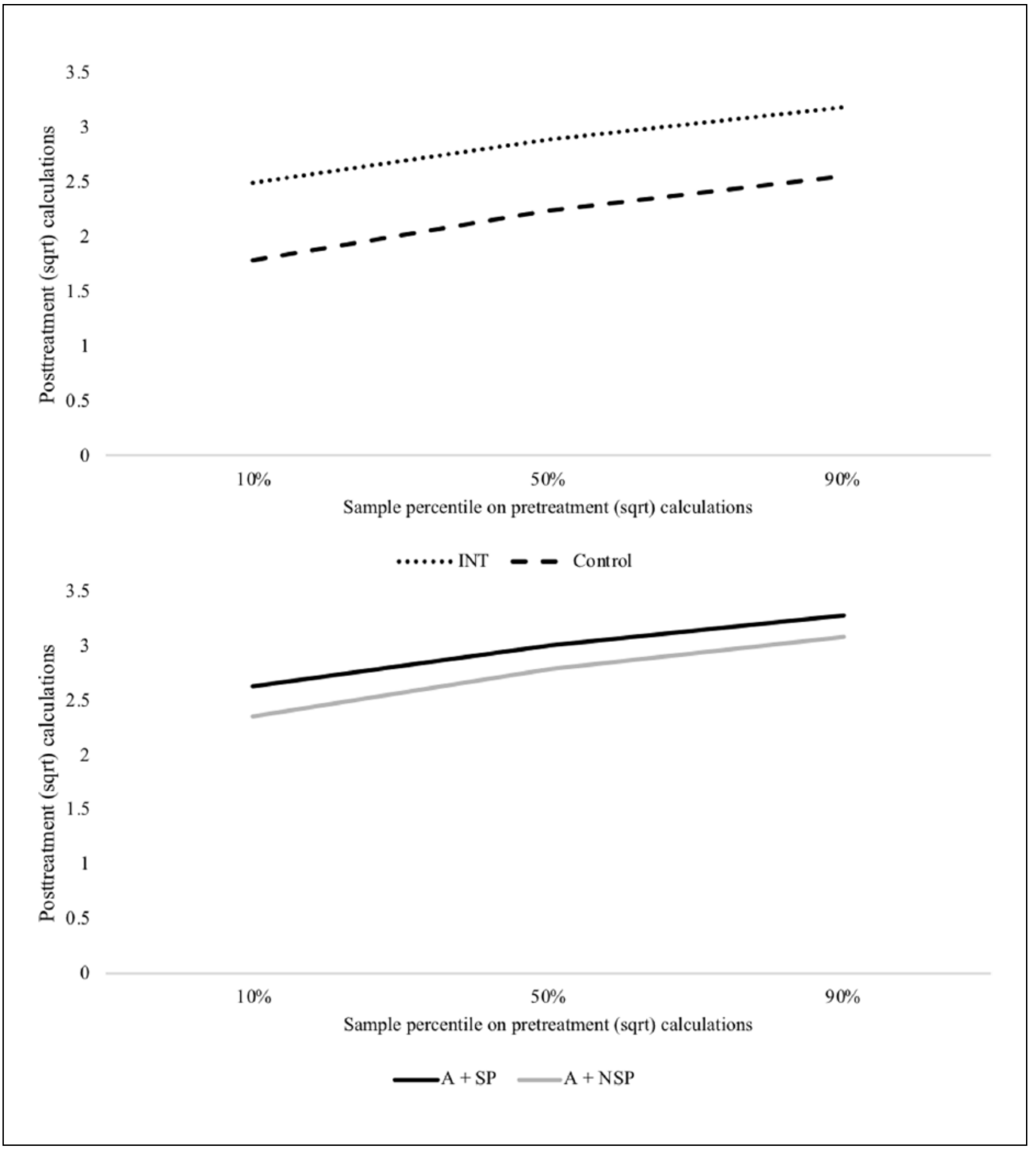

Figure 2. Graphical illustration of nonsignificant interaction effects on (square root) transfer to calculations: Top panel is intervention versus control conditions; bottom panel is A+SP (arithmetic concepts plus speeded practice) versus A+NSP (arithmetic concepts plus nonspeeded) conditions.

post-intervention performance level occurred for 71% of control students, 41% of A+SP students, and 46% of A+NSP students. On the number knowledge transfer outcome, inadequate growth occurred for 55% of control students, 23% of A+SP students, and 27% of A+NSP students. By contrast, inadequate post-intervention performance level occurred for 84% of control students, 68% of A+SP students, and 74% of A+NSP students.

Discussion

In this discussion, we interpret main effect results favoring intervention (across both practice conditions) versus control on all three outcomes and favoring speeded over nonspeeded practice on arithmetic. Then, we consider implications of finding that students' pre-intervention math skill did not moderate these main effects. Finally, we use individual

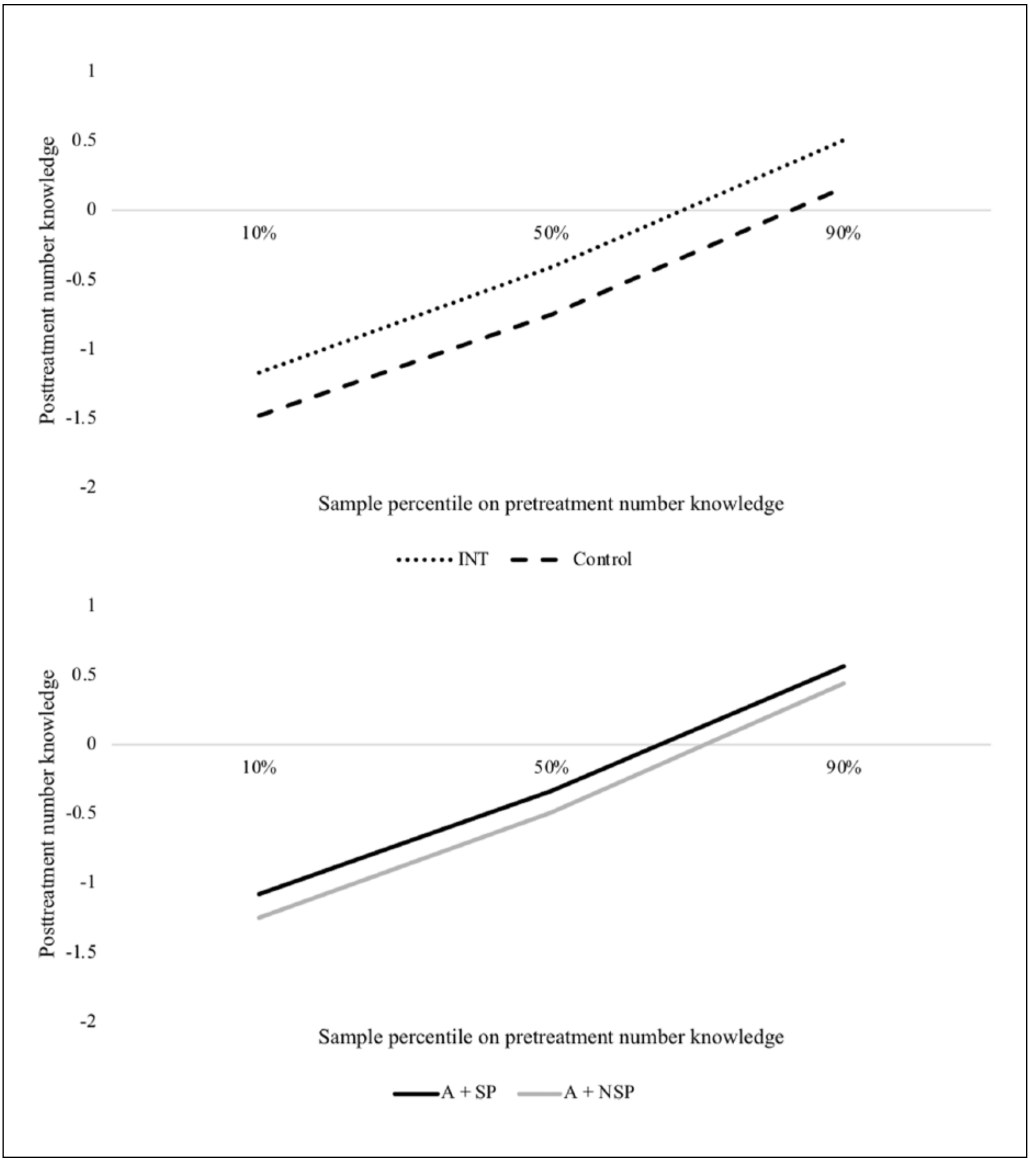

Figure 3. Graphical illustration of nonsignificant interaction effects on number sense transfer: Top panel is intervention versus control conditions; bottom panel is A+SP (arithmetic concepts plus speeded practice) versus A+NSP (arithmetic concepts plus nonspeeded) conditions.

student responsiveness data to contextualize results.

Effects of Number Knowledge Intervention Compared to the At-Risk Control Group

Students who received intervention significantly outperformed at-risk control group children on all three outcomes. This was the case even on this study's most distal outcome, the Number Sets Test (Geary et al., 2009). This task was novel to intervention students, whereas in some first-grade efficacy studies, number knowledge is indexed using tasks aligned with activities explicitly taught and practiced during intervention, such as magnitude comparison, counting, and ordering. With the Number Sets Test, by contrast, children are challenged with an integrative task in which they combine the sides of dominos that show arrays of objects and Arabic numerals to determine whether sums match target numbers. The task is speeded to simultaneously index accuracy on and fluency with cardinality, subitizing, counting, identifying numerals, understanding symbolic and nonsymbolic quantity, number decomposition, and arithmetic principles. In this study, this number knowledge measure represented a transfer challenge. So the ES of 0.27, favoring intervention over at-risk control group students, is notable.

Across the three outcomes, the main effects contrasting both intervention conditions against the at-risk control group provide persuasive evidence of the efficacy of this mathematics intervention.

In terms of the study's near transfer task, two-digit calculations with and without regrouping, intervention introduced students to place value and calculation strategies, but this focus was limited to 6 of 48 sessions. Therefore, the ES favoring intervention over at-risk control of 0.53 was impressive and probably carried at least in part by intervention students' superior arithmetic skill. On arithmetic, which was the central focus of intervention, the ES was large: 0.69. This is important because arithmetic is a core mathematical competency and a critical intervention target for first-grade children at risk for mathematics learning disabilities (e.g., Fuchs et al., 2006; Geary, 2011; Geary et al., 2012; U.S. Department of Education, 2008; Goldman et al., 1988; Jordan et al., 2003). Across the three outcomes, the main effects contrasting both intervention conditions against the at-risk control group provide persuasive evidence of the efficacy of this mathematics intervention.

Does Speeded Practice Provide Added Value Over Nonspeeded Practice?

The Fuchs et al. (2013) RCT also compared the effects of two practice conditions, each conducted in the context of arithmetic concepts instruction. Nonspeeded practice encouraged application of a variety of number-principle strategies; speeded practice encouraged efficient counting strategies. On arithmetic, results clearly favored speeded practice. The advantage of speeded over nonspeeded practice was associated with an ES of 0.44, and as reported in Fuchs et al. (2013), the estimate specifically for number knowledge intervention with speeded practice over the at-risk control group was 0.87 (vs. 0.51 for number knowledge intervention with nonspeeded practice over control). Results thus indicate a substantial contribution for speeded strategic practice in improving arithmetic outcomes.

In interpreting this finding, readers should note that schools sometimes provide timed practice without sufficient scaffolding in arithmetic concepts and in massed doses without support for immediate corrections of errors. This study's intervention, by contrast, delivered speeded practice in the context of rich, multifaceted instruction on the conceptual bases for arithmetic and formulated practice on a distributed basis to help children generate many correct responses, develop fluency with efficient counting strategies, immediately correct errors, and engage in strategic metacognitive behavior (i.e., retrieving answers from memory, if confident, otherwise using an efficient counting strategy). Therefore, findings generalize only to speeded practice that incorporates similarly sound, theoretically motivated instructional design.

Does Pre-Intervention Math Performance Moderate Intervention Effects?

A well-designed and executed RCT is the gold standard for validating an intervention. Yet, as noted in the introduction to this article, *validation* does not mean all students respond, and little is known about which student characteristics are associated with inadequate response. With the present analysis, we considered whether the severity of students' pre-intervention math deficits moderates the effects of generally effective first-grade arithmetic intervention.

We hypothesized that students with weaker pre-intervention performance profit less from intervention because severely discrepant initial performance may signify a severe form of learning difficulty (i.e., a learning disability), requiring a more individualized or sustained (i.e., intensive) form of intervention. This would be revealed in a moderator effect in which the learning advantage for the intervention over the at-risk control group is weaker for students with lower pre-intervention math skill than students with higher pre-intervention math skill. Understanding the tenability of this hypothesis is important for gauging the robustness with which an intervention addresses the full range of at-risk learners and identifying interventions that do and do not adequately address the needs of at-risk students with severely low pre-intervention math skill.

Our analyses looked at a pre-intervention math skill moderator effect on each outcome for two intervention contrasts: the effects of intervention (a) between students who receive the number knowledge intervention (aggregated across both practice conditions) versus at-risk control children and (b) between students in the two practice conditions. Contrary to expectations, for both contrasts on all three outcomes, the effects of intervention operated in parallel ways, regardless of the level of students' pre-intervention math skill.

Thus, to the major question posed in this article, we conclude that the effects of first-grade arithmetic intervention versus the at-risk control group on all three outcomes are robust across the distribution of children's pre-intervention mathematics performance. This is also the case for the superiority of speeded practice over non-speeded practice on arithmetic. So, severely low-performing students, presumed to have most severe risk for learning disabilities, benefit from conceptual arithmetic intervention comparably well, and they also enjoy stronger outcomes on arithmetic when intervention focused on the conceptual bases for arithmetic is combined with speeded strategic practice.

[T]he effects of first-grade arithmetic intervention versus the at-risk control group on all three outcomes are robust across the distribution of children's pre-intervention mathematics performance.

This finding corroborates Fuchs et al. (2016), in which a fractions intervention proved similarly efficacious across the continuum of pre-intervention whole-number performance. This suggests that mathematics interventions are robust across pre-intervention mathematics skill levels, although this question needs to be investigated on an intervention-specific basis. For the intervention at hand, *Galaxy Math*, finding that pre-intervention mathematics skill is *not* an indicator of inadequate response to intervention still leaves open the search for individual differences that are associated with responsiveness to intervention. Such information is needed to help schools circumvent students experiencing months of failure to an intervention that will eventually prove inadequate.

Rates of Individual Student Responsiveness: Implications for Intensive Intervention

The present study's analyses of rates of individual student responsiveness make clear why the search for individual differences associated with responsiveness to intervention is important. On the study's aligned outcome, arithmetic, the response rate, when indexed in terms of normalized growth, was strong: Only 6% of intervention students in the speeded practice condition and 13% of intervention students in the nonspeeded practice condition responded inadequately. In an absolute sense, these rates are encouragingly low, representing 1.5% to 3.25% of the general population. Even so, not all intervention children met the criterion for adequate improvement.

Moreover, when using normalized post-intervention performance level as the criterion, the inadequate response rate on the arithmetic outcome was higher: 48% of intervention students in the speeded practice condition completed intervention below the 25th percentile of not-at-risk classmates; 64% of intervention students in the nonspeeded practice condition. These rates of inadequate post-intervention performance among children who received the highly efficacious intervention are disturbingly high.

Higher rates of inadequate response when using post-intervention performance level than when using pre- to postintervention improvement have been reported elsewhere (e.g., Frijters et al., 2013; Fuchs et al., 2016). They suggest that although the vast majority of at-risk students improve nicely with highly efficacious intervention, many complete interventions inadequately prepared at-risk students to keep pace with classmates as new mathematics content was introduced in classrooms. This is not surprising given that at-risk students start intervention performing lower than not-at-risk classmates and because not-at-risk classmates are enjoying a period of rapid development on the three math outcomes during first grade.

At the same time, we caution readers that the criterion applied in the present study and commonly used in other responsiveness to intervention studies to determine inadequate post-intervention performance (scoring below the 25th percentile of a normative sample, here not-at-risk classmates) is arbitrary. Also, benchmarks for satisfactory post-intervention performance may need to differ as a function of grade level and the subdomain of mathematics in order to forecast long-term success in the general education program.

Therefore, a line of research is needed to provide the field with empirical standards for benchmark performance that distinguish students who should exit intervention from those who require more sustained or individualized (i.e., intensive) intervention. Requiring students to meet empirically derived post-intervention benchmarks prior to exiting intervention may address intervention fade-out effects (e.g., Clarke, Doable, Smolkowski, Nelson, Fien, Baker, & Kosty, 2016; Smith et al., 2013), in which validated interventions improve outcomes relative to an at-risk control group, as indexed at the end of intervention, but fail to decrease achievement gaps sufficiently to protect intervention at-risk children from mathematics difficulties in the long term.

We also note that in the present study, a similar pattern occurred on the two transfer outcomes, where higher rates of inadequate response occurred when the index was normalized post-intervention performance level than when normalized growth rate was employed. Rates of inadequate response (post-intervention performance level as well as on growth) were also higher on the complex calculations transfer outcome than for arithmetic. Inadequate response rates were higher still for the more integrative, challenging form of transfer on number knowledge. This underscores the importance of establishing empirical benchmarks for long-term success using a variety of outcomes, not just those proximal to the intervention content.

Conclusions

This analysis provides the basis for four main conclusions. First, efficacy for the first-grade mathematics intervention, when conventionally framed as stronger outcomes for at-risk intervention students compared to at-risk control students, is strong. Without intervention, students complete first grade with demonstrably and reliably worse mathematics performance on arithmetic, complex calculations, and integrative number knowledge than would be the case without that intervention. Second, caution is in order, as revealed in the individual student response data. These analyses remind us that strong efficacy does not provide the basis for assuming all at-risk students respond. This indicates the importance of identifying reliable methods to forecast, *before* intervention begins, which students need to proceed directly to more intensive intervention.

[R]esearch is needed to provide schools with technically strong post-intervention benchmarks for identifying students who are not adequately prepared to exit intervention and instead require more sustained, intensive services to avoid long-term failure.

In the present analyses, however, students' pre-intervention math performance did not provide the basis for forecasting which students will and will not demonstrate adequate response. This was also the case for a fourth-grade fractions intervention (Fuchs et al., 2016). Our third conclusion, therefore, is that intervention efficacy researchers must continue to examine moderator effects, considering not only pre-intervention math skill but also other student-level variables theoretically connected to the design of interventions.

In the absence of a reliable means for forecasting which students will and will not respond adequately to an intervention, the focus turns to methods for reliably distinguishing students who have and have not adequately responded at the end of intervention. Our fourth conclusion is that research is needed to provide schools with technically strong post-intervention benchmarks for identifying students who are not adequately prepared to exit intervention and instead require more sustained, intensive services to avoid long-term failure.

It is therefore unfortunate that few reports of intervention efficacy contextualize at-risk intervention student outcomes with respect to not-at-risk classmates (or with respect to normative frameworks from nationally normed tests). The goal of a multitier system of supports is to provide short-term intervention in a timely way to provide at-risk students with an academic boost that helps them succeed in the general education program without further support. Addressing the needs of approximately 50% of at-risk learners with validated intervention, as in the present analysis (across arithmetic and transfer measures), is an important contribution toward this goal. Equally important is timely identification of the remaining students who require intensive (more sustained or more individualized) intervention.

References

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B.*, *57*, 289–300. doi:10.2307/2346101

Clarke, B., Doable, C., Smolkowski, K., Nelson, E. K., Fien, H., Baker, S. K., & Kosty, D. (2016). Testing the immediate and long-term efficacy of a tier 2 kindergarten mathematics intervention. *Journal of Research on Educational Effectiveness*, *9*, 607–634. doi:10.1080/19345747.2015.1116034

Frijters, J. C., Lovett, M. W., Sevcik, R. A., & Morris, R. D. (2013). Four methods of identifying change in the context of a multiple component reading intervention for struggling middle school readers. *Reading and Writing: A Contemporary Journal*, *26*, 539–563. doi:10.1007/s11145-012-9418-z

Fuchs, L. S. (2003). Assessing treatment responsiveness: Conceptual and technical issues. *Learning Disabilities Research and Practice*, *18*, 172–186. doi:10.1111/1540-5826.00073

Fuchs, L. S., Fuchs, D., & Bryant, V. J. (2010). *Galaxy math*. Nashville, TN: Vanderbilt University.

Fuchs, L. S., Hamlett, C. L., & Powell, S. R. (2003a). *First-grade mathematics assessment battery*. Nashville, TN: Vanderbilt University

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Schatschneider, C, Hamlett, C., . . . Changas, P. (2013). Effects of first-grade number knowledge tutoring with contrasting forms of practice. *Journal of Educational Psychology*, *105*, 58–77. doi:10.1037/a0030127

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., . . . Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98*, 29–43. doi:10.1037/0022-0663.98.1.29

Fuchs, L. S., Sterba, S. K., Fuchs, D., & Malone, A. (2016). Does evidence-based fractions intervention address the needs of very low-performing students? *Journal of Research on Educational Effectiveness*, *9*, 662–677. doi:10.1080/19345747.2015.1123336

Fuson, K. C., & Kwon, Y. (1992). Learning addition and subtraction: Effects of number word and other cultural tools. In J. Bideau, C. Meljac, & J. P. Fisher (Eds.), *Pathways to number* (pp. 351–374). Hillsdale, NJ: Erlbaum.

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

Geary, D. C. (2011). Cognitive predictors of individual differences in achievement growth in mathematics: A five year longitudinal study.

Developmental Psychology, *47*, 1539–1552. doi:10.1037/a0025510

Geary, D. C., Bailey, D. H., & Hoard, M. K. (2009). Predicting mathematical achievement and mathematical learning disability with a simple screening tool: The Number Sets Test. *Journal of Psychoeducational Assessment*, *27*, 265–279. doi:10.1177/0734282908330592

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. In U.S. Department of Education (Ed.), *The final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.

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

Goldman, S. R., Pellegrino, J. W., & Mertz, D. L. (1988). Extended practice of addition facts: Strategy changes in learning-disabled students. *Cognition and Instruction*, *5*, 223–265. doi:10.1207/s1532690xci0503_2

Groen, G. J., & Resnick, L. B. (1977). Can preschool children invent addition algorithms? *Journal of Educational Psychology*, *69*, 645–652. doi:10.1037/0022-0663.69.6.645

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology*, *85*, 103–119. doi:10.1016/S0022-0965(03)00032-8

LeFevre, J. A., & Morris, J. (1999). More on the relation between division and multiplication in simple arithmetic: Evidence for mediation of division solutions via multiplication. *Memory & Cognition*, *27*, 803–812. doi:10.3758/BF03198533

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, & J. Cantler (Eds.), *Handbook of response to intervention (RTI) in early childhood education* (pp. 41–56). Baltimore: Brookes.

Rickard, T. C., Healy, A. F., & Bourne, L. E., Jr. (1994). On the cognitive structure of basic arithmetic skills: Operand order, symbol, and operation transfer effects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *20*, 1139–1153. doi:10.1037/0278-7393.20.5.1139

Siegler, R. S., & Robinson, M. (1982). The development of numerical understandings. In H. W. Reese & L. P. Lipsitt (Eds.), *Advances in child development and behavior* (pp. 242–312). New York, NY: Academic Press.

Smith, T., Cobb, P., Farran, D. C., Cordray, D. S., & Munter, C. (2013). Evaluating Math Recovery: Assessing the causal impact of a diagnostic tutoring program on student achievement. *American Educational Research Journal*, *50*, 397–428. doi:10.3102/0002831212469045

Sterba, S. (2017). Partially nested designs in psychotherapy trials: A review of modeling developments. *Psychotherapy Research*, *27*, 425–436. doi:10.1080/10503307.2015.1114688

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

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013). *What works clearinghouse: Procedures and standards handbook (version 3.0)*. Retrieved from http://whatworks.ed.gov/

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

Wilkinson, G. S. (1993). *Wide Range Achievement Test 3*. Wilmington, DE: Wide Range.

Authors' Note

This research was supported by Grant R324D130003 from the National Center on Special Education Research within the Institute on Education Sciences in the U.S. Department of Education to Vanderbilt University and by 2 R01 HD053714 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 National Center on Special Education Research, 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.

Manuscript received April 2018; accepted May 2018.