School Psychology Review,
2014, Volume 43, No. 2, pp. 160–177

Preliminary Evaluation of a Tier 2 Mathematics Intervention for First-Grade Students: Using a Theory of Change to Guide Formative Evaluation Activities

Ben Clarke, Christian T. Doabler, and Mari Strand Cary
University of Oregon

Derek Kosty
Oregon Research Institute

Scott Baker
Southern Methodist University

Hank Fien
University of Oregon

Keith Smolkowski
Oregon Research Institute

Abstract. This pilot study examined the efficacy of a Tier 2 first-grade mathematics intervention program targeting whole-number understanding for students at risk in mathematics. The study used a randomized block design. Students ($N =$ 89) were randomly assigned to treatment (Fusion) or control (standard district practice) conditions. Measures of mathematics achievement were collected at pretest and posttest. Treatment and control students did not differ on mathematics assessments at pretest. A series of random-effects models were estimated to compare gains between treatment and control conditions. Gain scores of intervention students were significantly greater than those of control peers on a proximal measure of mathematics achievement. The role of a strong theory-of-change model in the development and evaluation of mathematics interventions is articulated. Implications for researchers and educators designing and delivering instruction for at-risk students in a response-to-intervention model are discussed.

The research reported in this article was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R324A090341 to the Center on Teaching and Learning at the University of Oregon. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education.

Correspondence regarding this article should be addressed to Ben Clarke, Center on Teaching and Learning, University of Oregon, Eugene, OR 97403; e-mail: clarkeb@uoregon.edu

The critical importance of mathematics has garnered increased attention in the past decade (National Mathematics Advisory Panel [NMAP], 2008; National Research Council [NRC], 2001). The most recent National Assessment of Educational Progress (NAEP) results classified 58% of fourth-grade students as failing to reach proficiency in mathematics and 17% as falling below basic achievement patterns on the NAEP; the results are even more disconcerting when examined by income, ethnicity, and disability status (National Center for Education Statistics, 2013). Young students without a deep understanding of mathematics risk losing access to more advanced mathematics including algebra (NMAP, 2008) and long-term career opportunities available in the fields of science, technology, mathematics, and engineering (National Science Board, 2008). The National Council of Teachers of Mathematics (2013) noted that "an economically competitive society recognizes the importance of mathematics learning to adult numeracy and financial literacy, and it depends on citizens who are mathematically literate" (p. 1). With recognition of the negative impact of low mathematics achievement at both the individual and national level, urgent calls from the highest levels of the federal government have been made for an increased focus on improving the mathematics outcomes of our nation's students (Obama, 2013).

Occurring simultaneously with lower than desired levels of mathematics achievement is a growing recognition that a successful start in mathematics is critical in ensuring long-term success. Morgan, Farkas, and Wu (2009) analyzed longitudinal data from the Early Childhood Longitudinal Study database and found that of the students who entered and exited kindergarten below the 10th percentile, 70% remained below the 10th percentile in fifth grade. In contrast, of the students who entered kindergarten below the 10th percentile but exited above the 10th percentile, only 30% were below the 10th percentile in fifth grade. In other words, those students who came into kindergarten at an elevated risk for math difficulties but grew substantively over the course of the year were markedly less likely to be at risk up to 5 years later. These trends found in longitudinal data sets of mathematics achievement mirror those found for the development of reading trajectories (Juel, 1988). Such findings in the area of reading development spurred a focus on prevention of reading difficulties through the use of screening systems to identify at-risk students (Good, Gruba, & Kaminski, 2002) and the development of curriculum materials targeting foundational reading skills (Wanzek & Vaughn, 2010). A similar system, based on the idea of preventing mathematics difficulties before they fully develop by identifying at-risk students and providing early intervention services targeting key foundational skills, has been advocated in mathematics (Fuchs, Fuchs, & Compton, 2013).

The focus on prevention of mathematics difficulties fits within the context of service delivery in schools based on a tiered model of instruction commonly referred to as *response to intervention* (RtI; National Association of State Directors of Special Education, 2006). Though originally conceptualized as a procedure to evaluate eligibility for special education services (Individuals with Disabilities Education Improvement Act, 2004), in practice RtI has been implemented as a more robust system of support to increase the achievement of all students (Fuchs, Fuchs, & Zumeta, 2008; Vaughn & Fuchs, 2003). The shift in conceptualization has placed a tighter focus on the instructional supports provided to students at different levels of need, including the instruction provided as part of the core classroom experience (i.e., Tier 1) and additional instructional support (i.e., Tiers 2 and 3) provided to students who do not respond to research-based core instruction. RtI systems have, in some respects, become standard in reading (Vaughn, Wanzek, Woodruff, & Linan-Thompson, 2007), whereas in mathematics key RtI components require further investigation (Bryant et al., 2011) to meet the need for research- and evidence-based programs (Glover & DiPerna, 2007) vital to any RtI system.

Current Research on Tier 2 Mathematics Intervention Programs

Any call for improved mathematics achievement within an RtI model is dependent on educators being able to access state-of-the-art curricular programs designed to address the specific needs of students attempting to gain access to mathematics content (Clarke, Baker, & Chard, 2008). The current research base in mathematics, while expanding, lags behind the field's knowledge base regarding reading instruction (Gersten et al., 2009). We reviewed 78 elementary math programs that had been evaluated by the What Works Clearinghouse (WWC) and found that only 7 of them were evaluated by studies that met the WWC standards, with only 4 programs showing potential positive effects on student achievement. Two concerns are raised by this review. First, fewer than 10% of programs had been evaluated with research designs of sufficient rigor to enable conclusions to be drawn regarding their impact. Second, of the four programs showing a potentially positive impact, all four were core (Tier 1) programs and not designed specifically for use with at-risk students.

A second analysis found that only nine intervention studies had been conducted on programs suitable for use as Tier 2 programs in an RtI model (Newman-Gonchar, Clarke, & Gersten, 2009). Of those nine studies, only two were designed for use with first-grade students. In the first study (Fuchs et al., 2005), a randomized controlled trial design was used to test the efficacy of a 63-lesson program, Number Rockets, on mathematics achievement. In each lesson, students received 30 min of small-group instruction on 17 key number concepts and then 10 min of computer-based instruction focused on increasing procedural fluency on mathematics facts. The results indicated a significant impact on three major areas of mathematics understanding—(a) computation, (b) concepts and applications, and (c) story problems—with effect sizes ranging from 0.11 to 0.70. An impact was not found on student fact fluency performance. A subsequent large-scale replication study evaluating Number Rockets was conducted in four states (Rolfhus et al., 2009), with similar but more moderate results, with an effect size of 0.34 on the Test of Early Mathematics Ability, Third Edition (Ginsburg & Baroody, 2003). In the second of the studies reviewed (Newman-Gonchar, Clarke, & Gersten, 2012), Bryant, Bryant, Gersten, Scammacca, & Chavez (2008) used a less rigorous regression discontinuity design to examine the impact of a small-group intervention program targeting key mathematical concepts, as well as number concepts and relationships, such as base 10 and place value. On average, 64 lessons were completed across 18 weeks. The study did not find a significant impact on either a proximal or distal measure of student achievement. It should be noted that in both studies, the focus was on evaluating impact, and the authors did not address the role of potential mediators of student outcomes.

With recognition of the need for expanding the research base on effective mathematics instruction, a number of seminal documents (Gersten et al., 2009; NRC, 2001) including the NMAP (2008) report have explicitly called for the development and rigorous evaluation of mathematics curricula. Thus research on core programs (programs designed for and used at the whole-classroom level) and intervention programs (Tiers 2 and 3) provided in the early elementary grades is critical.

Theory-of-Change Models in Curriculum Development and Evaluation

Foundational to the development of research-based curricula are frameworks that link curriculum development efforts to an underlying theory of change. Clements (2007) noted that "developers must draw from existing research so that what is already known can be applied to the anticipated curriculum" (p. 37) and, in turn, developers must "structure and revise the nature and content of curricular components in accordance with models of children's thinking and learning in a domain" (p. 37).

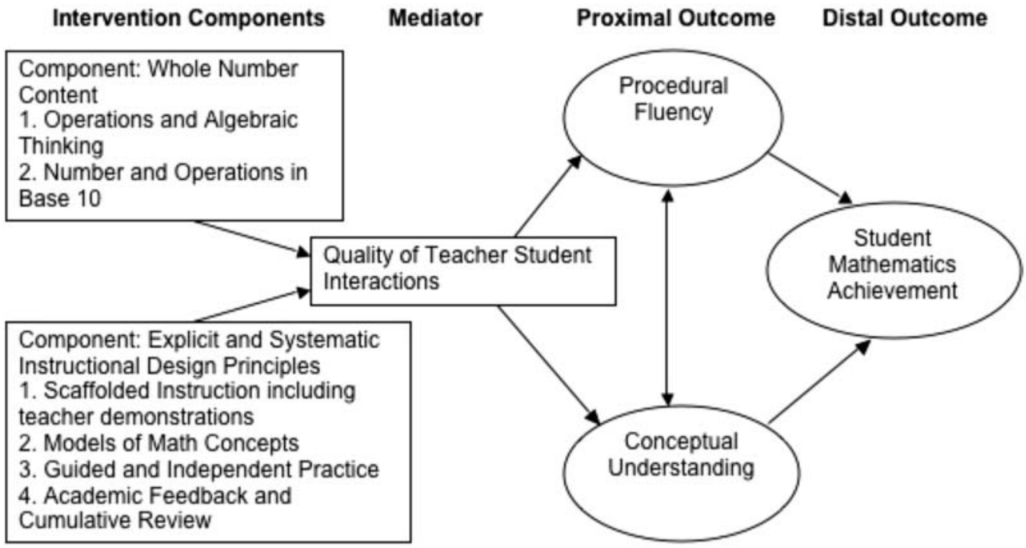

Figure 1. Fusion theory of change.

When using a theory of change, developers need to justify their predictions by drawing from the relevant theoretical and empirical knowledge bases. For example, researchers interested in developing an early mathematics curriculum will review existing research involving interventions for students with or at risk for mathematics difficulties. A strong theory of change also has roots in relevant theories of learning (e.g., Bransford & Donovan, 2005). By systematically grounding an intervention in the learning sciences, researchers are able to provide theoretical alignment between children's thinking and learning of mathematics and the instructional techniques embedded within an intervention (Clements, 2007). A strong theory of change also addresses the variables hypothesized to mediate and moderate the impact of an intervention. As Rothman (2013) observed, "Mediators and moderators are the building blocks of theory and, in turn, intervention design, specifying the connections between these two classes of constructs is at the heart of developing, testing, and refining theory" (p. 190). *Mediating variables* refer to the processes that comprise an intervention, whereas *moderating variables* are student and teacher factors that may potentially change the relationship between an intervention and student outcomes. Establishing mediators and moderators in a theory of change offers researchers the opportunity to unpack the "black box" of classroom instruction by ascertaining whether an intervention is more effective under certain conditions or with a particular subgroup of the student population (MacKinnon & Luecken, 2008).

Fusion's Theory of Change

Our efforts to develop and evaluate a research-based first-grade mathematics intervention curriculum, Fusion, were guided by an underlying theory-of-change model. As depicted in Figure 1, the theory of change for the Fusion intervention is composed of three key levels: (a) intervention components, (b) mediator variables, and (c) proximal and distal student outcomes. The Fusion intervention contains two key components: whole-number content and explicit and systematic instructional design principles. When carefully integrated, these intervention components are expected to facilitate instructional interactions between teachers and students around foundational whole-number concepts and skills. We hypothesize that the quality of these instructional interactions will mediate students' proximal outcomes (labeled *conceptual understanding* and *procedural fluency* in our theory of change). It is hypothesized that these two proximal outcomes will have a direct impact on student mathematics achievement, which is labeled as a *distal outcome* in Figure 1. In the following section, we describe each level of Fusion's theory of change and summarize the importance of each in the relevant literature.

Intervention Component: Whole-Number Content

Fusion's first intervention component attends to the calls from expert panels (NMAP, 2008; NRC, 2001, 2009) for early mathematics curricula to have greater focus on the critical aspects of whole numbers, often referred to as *number sense*. Proficiency with number sense is essential for students' overall academic success throughout public school and the opportunities they have for meaningful postsecondary experiences (Morgan et al., 2009; NRC, 2001). A growing body of evidence suggests that many children, particularly children from economically and educationally disadvantaged backgrounds, do not possess a firm number sense and thus struggle with making quantitative comparisons, manipulating numbers and their operations, and grasping the connection between mathematical concepts and numerical relationships (Gersten & Chard, 1999). Although the definition of *number sense* varies among educational researchers and mathematicians (Berch, 2005), there is general consensus that early number sense leads to the automatic use of foundational math skills, such as completing written calculations and solving applied problems (Gersten & Chard, 1999; NMAP, 2008; NRC, 2001). In first grade, foundational attributes of number sense identified in the Common Core State Standards (Common Core State Standards Initiative, 2010) include place-value concepts, number combinations, multidigit problems involving addition and subtraction, and word-problem solving.

Instructional Component: Explicit and Systematic Design Principles

The second intervention component of Fusion is the incorporation of *explicit* and *systematic* instructional design principles to support students' development of mathematical proficiency. A consistent finding of empirical research is that explicit mathematics instruction has significant value for students struggling with mathematics (Baker, Gersten, & Lee, 2002; Gersten et al., 2009). For example, in a meta-analysis of 41 studies involving students with math disabilities, Gersten et al. (2009) reported that explicit instruction had a substantively important positive effect (Hedges's $g = 1.22$) on student mathematics achievement. Explicit instruction is a structured delivery approach that promotes learning mastery in the foundational concepts and skills of early mathematics. According to experts in the field, an early mathematics curriculum is considered explicit when it supports teachers in (a) introducing new and complex math content through unambiguous explanations and demonstrations, (b) facilitating frequent opportunities for students to practice with important mathematics content, and (c) providing timely academic feedback to confirm correct student responses and address potential misconceptions (Archer & Hughes, 2010; Doabler et al., 2013; Gersten et al., 2009).

As with explicit design principles, research has also shown the importance of systematically designing mathematics instruction for students with difficulties in mathematics. Systematic design principles attend to the way academic information, such as math concepts and skills, is selected, prioritized, and organized within and across a curriculum's lessons (Coyne, Kame'enui, & Carnine, 2011). For instance, a systematically designed curriculum will judiciously interweave and appropriately match visual representations of mathematics, such as place-value blocks, with abstract symbols to illustrate solution methods for math problems. A growing body of research shows that this concrete–representational–abstract approach supports students in formulating "well-developed knowledge packages" (Ma, 1999, p. 113) of fundamental mathematics (Gersten et al., 2009; Witzel, Riccomini, & Schneider, 2008).

Mediating Variables

The instructional interactions that take place between teachers and students around critical mathematics content are a defining characteristic of effective classroom instruction and, we hypothesize, mediate student outcomes. Classrooms in which these critical teacher–student interactions occur at a higher rate would have greater student outcomes regarding critical early mathematics concepts. The purpose of such teacher–student interactions is to facilitate meaningful opportunities for students to interact with and practice important mathematical concepts, skills, and procedures. The frequency of practice opportunities has important implications for promoting students' success in early mathematics, and findings from recent research suggest that a critical format of student practice is mathematical verbalization (Doabler et al., 2013; Gersten et al., 2009). Verbalizations offer opportunities for students—both specific individuals and the group at large—to communicate their mathematical thinking and understanding. In the early grades, math verbalizations can be a critical mode of student responding because they allow all students the opportunity to learn and participate. For example, a teacher can facilitate an entire class of students in explaining their solution methods for solving a multidigit addition problem.

Proximal and Distal Outcomes

Mathematics proficiency is composed of two knowledge forms: conceptual understanding and procedural fluency (NRC, 2001). *Conceptual knowledge* refers to an understanding of the relationship between representations of math concepts and abstract symbols, whereas the latter knowledge form entails automaticity of math procedures (Wu, 1999). Educational research has consistently shown that at-risk learners have difficulty making a connection between these two knowledge forms (Gersten et al., 2009). Therefore, in our theory of change (Figure 1), conceptual understanding and procedural fluency represent two proximal outcomes targeted by the Fusion intervention. We hypothesize that Fusion will support students' development of these two knowledge forms concurrently. That is, as Fusion helps students build understanding of math concepts, it will increase their fluency in solving math problems through strategically planned opportunities for guided and independent practice, as well as cumulative review. Furthermore, we hypothesize that the reciprocal relationship between conceptual understanding and procedural fluency will have a direct impact on students' overall mathematics achievement.

Purpose

The primary purpose of this randomized controlled trial pilot study is to test the impact of a first-grade intervention program, Fusion, on the achievement of students at risk in mathematics. There is an intensive need for rigorous efficacy trials of first-grade mathematics interventions as evidenced by the paucity of current research in the area (Gersten et al., 2009; WWC, 2013) and by calls for focused efforts on the development of intervention programs (Gersten et al., 2009; NMAP, 2008). We hypothesize that students in the Fusion condition will have greater student achievement outcomes. In addition, given that previous studies have focused exclusively on student outcomes, a secondary purpose is to begin exploring the underlying mechanisms that guide the design of intervention programs and potentially mediate the impact of intervention programs. A direct examination of mediation specifically requires showing that a given condition accounts for differences in implementation across conditions.

In this study, we were unable to conduct mediation analysis because we did not have implementation data from control classrooms. To attempt to navigate this barrier, we examined associations between implementation quality and student achievement gains within treatment classrooms. Because the Fusion program is scripted to ensure high degrees of critical teacher and student behaviors, we hypothesize that higher levels of implementation will result in a greater degree of teacher–student interactions and greater student outcomes. That is, teachers who teach with an overall level of high quality and implement the program with high levels of fidelity will be engaging with their students in the types of behavior hypothesized to mediate student achievement in our theory of change. For example, Fusion contains language that prompts teachers to ask individual questions of participating students; thus, if a teacher is implementing Fusion with a high level of fidelity, we expect to see more frequent teacher–student interactions as students respond to individual questions. The work in this study stands to contribute to the limited body of research on first-grade mathematics interventions, and by examining results within the context of a theory-of-change model, findings from the study may contribute to the growing knowledge base on effective mathematics instruction within an RtI model of system delivery.

Method

Participants

The study took place in nine schools with approximately 10 eligible students per school, based on screening scores and teacher recommendations. The research team randomly assigned these 10 students to intervention (Fusion instruction) or a control (standard district practice) by using a random number generator and assigning the lowest five to intervention. The final sample included 89 students: 44 in the intervention group and 45 in the control group. Control students did not receive Fusion instruction but were not prohibited from receiving standard district intervention services. All participants received standard classroom mathematics instruction.

Schools. The schools were drawn from two suburban school districts in the Pacific Northwest. District A (five schools) had 10,796 students: 33% were minorities, 6% were English-language learners, 60% were eligible for free/reduced-price lunch, and 15% received special education services. District B (four schools) had 5,866 students: 28% were minorities, 3% were English-language learners, 55% were eligible for free/reduced-price lunch, and 17% received special education services. The schools were from research partner districts in an Institute of Education Sciences development grant. District staff recruited schools within their district interested in participating.

Students. All first-grade students completed group-administered versions of the Quantity Discrimination (QD) and Missing Number (MN) measures. The group-administered QD and MN measures were modified versions of individually administered QD and MN measures (Clarke & Shinn, 2004). Raw scores on the screener were converted to *z* scores and averaged. The 10 lowest scoring students on the screener per school not meeting the exclusion criteria were identified and eligible for the study. We excluded students if they could not identify or write numbers 1 to 10 or if they had severely limited English proficiency (based on the judgment of the student's primary teacher). Demographic information for the sample is shown in Table 1.

Interventionists. Nine district employees (i.e., interventionists) taught one small Fusion group each. Interventionists were included in the study based on time and schedule availability. All of the interventionists were women. One was a high school graduate, two had bachelor's degrees, and six had master's degrees. On average, the interventionists had 8.7 years' teaching experience (range, 3–25 years), 7.4 years' experience teaching math (range, 3–25 years), and 7.7 years' experience teaching first grade (range, 4–20 years).

Measures

Fidelity of implementation. Each Fusion lesson consisted of at least three primary activities. Observers rated implementation fidelity for the first three primary activities in a Fusion lesson using a 0–1 scale (0, *not taught*; 0.5, *partial implementation*; and 1, *full imple-*

Table 1
Descriptive Information for Demographic Characteristics by Condition

Demographic Characteristic	Treatment (n = 44)	Control (n = 45)	χ^2 (1 df)	p Value
Male [n (%)]	21 (47.7)	29 (64.4)	2.53	.112
Nonwhite [n (%)]	4 (9.1)	9 (20.0)	2.12	.145
Hispanic [n (%)]	6 (13.6)	12 (26.7)	2.34	.126
Free/reduced-price lunch [n (%)]	31 (70.5)	31 (68.9)	0.03	.872
English-language learner [n (%)]	6 (13.6)	10 (22.2)	1.11	.292
Eligible for special education services [n (%)]	13 (29.5)	11 (24.4)	0.29	.588

mentation). A fidelity score for each observation was calculated by averaging ratings across Activities 1 through 3. Each interventionist's fidelity scores were averaged across the three observation occasions. Observers also provided a holistic rating of overall level of implementation on a 7-point scale, with a score of 1 representing low implementation and 7 representing high implementation.

Ratings of Classroom Management and Instructional Support. Ratings of Classroom Management and Instructional Support (RCMIS; Doabler & Nelson-Walker, 2009) is a holistic rating system composed of 14 items (e.g., clear and consistent delivery of instruction) that measure the quality of instructional interactions that take place between teachers and students around critical mathematics content (Cronbach's α = 0.92). Each curriculum-independent item is rated on a 4-point scale from *low* (1) to *high* (4). For each observation, a score was calculated by averaging the ratings across the 14 items. For each group, an overall quality score was calculated as the mean across all observations. The RCMIS was used as a measure of overall instructional quality.

Early Numeracy Curriculum-Based Measures. Early Numeracy Curriculum-Based Measures (EN-CBM; Clarke & Shinn, 2004) was used as a proximal measure of students' procedural fluency. All measures were timed for 1 min. The Oral Counting measure requires students to orally rote count as high as possible without making an error. Concurrent and predictive validities range from 0.46 to 0.72. For all EN-CBM measures, the criterion measures were the Number Knowledge Test (Okamoto & Case, 1996), Woodcock-Johnson Applied Problems subtest (Woodcock & Johnson, 1989), and Mathematics-CBM (Shinn, 1989). The predictive-validity timeframe was from the fall to the spring. The Number Identification measure requires students to orally identify numbers between 0 and 10 when presented with a set of printed number symbols. Concurrent and predictive validities range from 0.62 to 0.65. The QD measure requires students to name which of two visually presented numbers between 0 and 10 is greater. Concurrent and predictive validities range from 0.64 to 0.72. The MN measure requires students to name the missing number from a string of numbers (0–10). Concurrent and predictive validities range from 0.46 to 0.63. A total EN-CBM score, calculated by summing raw scores from the four subtests, was used in the analysis. Preliminary evidence indicates the measures' capability to monitor growth (Clarke & Shinn, 2004; Clarke et al., 2008).

Group curriculum-based measure. Two of the individually administered EN-CBM (Clarke & Shinn, 2004) measures, QD and MN, were adapted for small-group administration and used as a screening instrument. Whereas the original measures require students to verbally respond to each item, the group curriculum-based measure has them write their responses (circling the correct choice or filling in the missing number). The test–retest reliabilities of the group-administered QD and MN measures are 0.87 and 0.85, respectively. Concurrent and predictive validities with the ProFusion assessment range from 0.58 to 0.80 for the QD measure and from 0.42 to 0.57 for the MN measure (Doabler et al., in press).

Stanford Achievement Test, 10th Edition. The Stanford Achievement Test, 10th Edition (SAT-10; Harcourt Educational Measurement, 2002), is a group-administered, norm-referenced examination for kindergarten through twelfth-grade students. Two math subtests were used as distal measures of mathematics performance. The Math Problem Solving subtest assesses problem solving and mathematical reasoning. The Math Procedures subtest assesses computational fluency. The SAT-10 is a standardized achievement test with reliability estimates that exceed 0.90 and a criterion-related validity coefficient of approximately 0.60 to 0.70 (Harcourt Educational Measurement, 2002).

ProFusion. The ProFusion measure was developed by the research team to assess students' conceptual and procedural knowledge of number and numeration, place-value concepts, basic number combinations, and problems involving multidigit addition and subtraction. In an untimed, group setting, students are asked write numbers from dictation (four items) and numbers missing from a sequence (three items), write numbers matching base-10 block models (three items), and decompose double-digit numbers (three items). Moreover, students complete addition problems and subtraction problems (eight items) and story problems (two items). Students also complete 1-min, timed addition (32 items possible) and subtraction (24 items possible) fluency measures and work with proctors individually to complete the number-identification portion (8 items). The criterion validity with other posttest measures used in the study was $r = 0.56$ with the EN-CBM total score and $r = 0.68$ with the SAT-10.

The measurement net for the study was designed to represent the theory of change outlined in the introduction section. The ProFusion measure functioned as a measure of proximal conceptual understanding, and EN-CBM measures were selected to function as a proximal measure of procedural fluency. The SAT-10 measure was used as a distal measure of mathematics achievement. The RCMIS and fidelity-of-implementation measures were used to examine overall instructional quality and teacher–student interactions.

Procedures

Data collection. Prior to beginning data collection, data collectors with experience in conducting educational assessments for research projects attended 2 days of training. Data collectors for the EN-CBM measures and the SAT-10 were not affiliated with the project in any other manner (e.g., interventionists, authors of this article). Fusion interventionists administered the ProFusion assessment after a half day of training. During training on individually administered assessments, data collectors were shadow scored on a practice administration and held to a 90% interscorer reliability standard. A fidelity checklist (e.g., reads directions as standardized) was used for all measures to ensure reliable administration. Similar procedures were followed during data collection in the field, with shadow scoring to a criterion of 90% interscorer reliability on individually administered measures and the use of fidelity checklists on all measures. Once data were collected, all protocols were double scored and double entered by two data collectors. All first-grade students completed the group QD and MN screeners approximately 1 month before the start of the intervention. Participating students completed the EN-CBM, SAT-10, and ProFusion at pretest before the start of Fusion instruction in their schools. After Fusion instruction ended, participants completed the EN-CBM, SAT-10, and ProFusion at posttest. Pretest data were collected in the 2 weeks before the start of the intervention, and posttest data were collected during a 2-week window after the intervention was completed.

Trained project staff observed each group's Fusion instruction three times (i.e., once during the beginning, middle, and end of the curriculum). Observers completed the RCMIS during the same observations as the fidelity-of-implementation instrument. RCMIS interobserver reliability assessment was conducted on 20% of all observation occasions, and RCMIS interobserver reliability was 91%. The reliability of the RCMIS was calculated by summing the total points by each observer and then dividing the smaller sum by the larger sum. Fidelity-of-implementation interobserver reliability assessment was conducted on 20% of all observations. Exact agreement (e.g., 100% reliability was found when both observers scored an activity with the same rating and 0% reliability indicated different ratings) was used to calculate reliability. Interobserver reliability was 95% and 86% for the activity-based rating and holistic rating, respectively.

Fusion intervention. The Fusion curriculum is a Tier 2 Grade 1 mathematics intervention designed for students at risk in whole-number concepts and skills. Students are taught in small groups of approximately five students and receive 60 lessons, each lasting 30 min, delivered over a period of 20 weeks. In this study, on average, the intervention lasted 18.6 weeks and teachers delivered 50.7 lessons.

Each lesson includes the explicit introduction of new content and systematic practice and review in four to five brief, scripted mathematics activities. Lessons use a variety of math models and contain teacher modeling, scaffolded instructional examples, and opportunities for teachers to provide academic feedback based on student responses to individual and group questions. Two mathematical domains in the first-grade Common Core State Standards—Operations and Algebraic Thinking and Number and Operations in Base Ten—form the basis of Fusion content. The first half of the curriculum emphasizes number sense, basic number combinations, and place-value concepts. During the second half of the curriculum, students encounter multidigit computation without regrouping and word-problem solving. In this study, interventionists were given guidelines to deliver one lesson per day, three times per week, in small-group instructional formats, with approximately five students per group.

Professional development. Interventionists participated in two 3-hr professional development workshops led by the authoring and research team. Workshops were intended to deepen content knowledge for teaching mathematics, pedagogical knowledge, and comfort teaching Fusion lessons. Workshops provided time to practice teaching Fusion lessons and receive feedback from the interventionists' peers and the curriculum's authors. The first workshop was conducted approximately 1 month before Fusion instruction. Content included an overview of the study design and the interventionists' role, an overview of the Fusion intervention and its underlying principles and content, lesson demonstrations, group management tips, and practice opportunities. The second training workshop occurred after interventionists had implemented approximately one quarter of the Fusion lessons. During this training, interventionists had the opportunity to ask questions about the first half of the curriculum and were introduced to concepts in the second half of the curriculum. There was no set standard that interventionists were required to meet during training prior to implementation.

Statistical Analysis

A series of random-effects models were estimated using the SPSS MIXED procedure to compare gains in ProFusion, EN-CBM, and SAT-10 outcomes between the treatment and control conditions. Raw scores were used in the analysis. The random-effects models nested pretest and posttest assessments within students and students within instructional groups. The models included the effects of time (coded 0 for pretest and 1 for posttest), condition of instructional group (coded 0 for control and 1 for treatment), and the Condition-by-Time interaction. The Condition-by-Time interaction represents the difference in gains in outcomes between the two groups. Hedges's g was used as a metric of intervention effect size for each outcome (0.2, 0.5, and 0.8 are considered small, medium, and large effects, respectively; WWC, 2011). As recommended by Feingold (2009), Hedges's g was computed as the Condition-by-Time interaction effect divided by the posttest pooled standard deviation of the outcome. In accordance with an intent-to-treat approach, maximum likelihood estimation was used to obtain model parameters and standard errors using all cases available, which results in less bias in parameter estimates and standard errors than other methods of handling missing data (e.g., listwise deletion; Schafer & Graham, 2002).

A second set of analyses was conducted to examine the relationships between (a) gains in student outcomes from pretest to posttest and (b) fidelity of Fusion implementation across Activities 1 through 3 and quality-of-instruction (RCMIS) ratings averaged across the three observation occasions. These analyses involved a series of random-effects models nesting students assigned to the treatment condition within Fusion instructional groups. Associations were estimated by regressing gain scores for each outcome on each observation measure separately. We report standardized parameter estimates (Snijders & Bosker, 1999).

Results

Baseline Equivalency and Attrition

The expectation of baseline equivalency owing to random assignment of groups was examined. The treatment and control groups were compared regarding demographic characteristics and outcome measures collected at pretest. Contingency-table analyses and t tests were conducted on categorical and continuous measures, respectively. The groups did not significantly differ on any demographic characteristics or pretest outcome measures (Table 1). The extent to which attrition threatened the internal and external validity of the study was evaluated using contingency-table analyses and analysis of variance. Participants who completed all posttest assessments were compared with those who did not with respect to demographic characteristics and pretest outcome measures. We also conducted 2-way analyses of variance to test whether outcome variables were differentially affected across conditions by attrition. These latter analyses examined the effects of condition and attrition status, as well as their interaction, on pretest outcomes. Among the 45 students assigned to the Fusion condition and the 44 control students, the attrition rates were 13.3% ($n = 6$) and 13.6% ($n = 6$), respectively. The attrition rates did not significantly differ by condition. We found no statistically significant differences in demographic characteristics or baseline outcomes by attrition status nor did we find any statistically significant interactions between attrition and condition predicting baseline outcomes, suggesting that attrition was not systematic.

Intervention Effects for Fusion

Table 2 provides descriptive statistics and intervention effects for each outcome measure. The treatment group had statistically significantly greater gains on our proximal measure of conceptual understanding, ProFusion, compared with control participants (estimate = 12.9, $p = .015$, Hedges's $g = 0.82$), corresponding to a large effect. The difference between groups was not statistically significant with respect to gains on our proximal measure of procedural fluency, EN-CBM (estimate = 7.8, $p = .667$, Hedges's $g = 0.14$), or on scores on our distal measure of conceptual understanding, SAT-10 (estimate = 1.1, $p = .590$, Hedges's $g = 0.11$).

Fidelity of Implementation, Quality of Instruction, and Student Performance

Table 3 provides descriptive statistics for the fidelity of implementation across Activities 1 through 3 and quality-of-instruction ratings averaged across the three observation occasions. To serve as a proxy for our hypothesized mediator, teacher–student interactions, associations between these measures and gains in student outcomes are also summarized in Table 3, which summarizes standardized pa-

Table 2
Pretest and Posttest Descriptive Statistics and Condition-by-Time Intervention Effects for Outcome Measures

	Pretest			Posttest			Condition-by-Time Intervention Effect			
Measure	*M*	*SD*	*n*	*M*	*SD*	*n*	Estimate	*t*	*p*	Hedges's *g*
ProFusion							12.9	2.71	.015	0.82
Fusion	23.0	11.2	44	53.1	15.1	38				
Control	27.2	13.4	42	44.0	16.3	39				
EN-CBM							7.8	0.44	.667	0.14
Fusion	143.0	42.1	44	183.2	61.8	40				
Control	148.8	45.9	44	182.2	52.1	41				
SAT-10							1.1	0.55	.590	0.11
Fusion	22.6	6.5	44	33.4	10.4	38				
Control	23.2	7.3	43	32.5	9.9	40				

Note. Tests of the Condition-by-Time interaction used 16 *df*. EN-CBM = Early Numeracy Curriculum-Based Measure; SAT-10 = Stanford Achievement Test, 10th Edition.

rameter estimates (βs). Although none of the relationships were statistically significant ($p > .15$ for all tests), moderate to large positive associations were found between gains in (a) ProFusion outcomes and implementation of Activities 1 through 3 (β = 0.21), overall implementation fidelity (β = –0.20), teachers providing models (β =

Table 3
Descriptive Statistics for Fidelity of Implementation and Quality of Instruction Ratings and Their Associations With Student Outcomes

			Associations With Student Outcome (β)		
Measure	*M* (*SD*)	Range	ProFusion	EN-CBM	SAT-10
Fidelity of implementation[a]					
Implemented Activities 1 through 3	0.9 (0.1)	0.8–1.0	0.21	–0.39	0.04
Modeled skill or concept	0.9 (0.2)	0.6–1.0	0.50	–0.23	0.26
Provided group response opportunities	0.9 (0.2)	0.5–1.0	–0.05	0.13	–0.21
Provided individual turns	0.9 (0.1)	0.7–1.0	0.32	–0.23	0.16
Provided academic feedback	0.9 (0.1)	0.6–1.0	–0.08	0.04	–0.19
Overall[b]	5.2 (1.1)	3.3–6.3	–0.03	–0.20	–0.16
Quality of instruction[c]	3.2 (0.6)	2.4–3.7	–0.04	–0.09	–0.26

Note. Items and summary scores were averaged across three observation occasions. Tests of associations (fixed effects) used 7 *df*; with $p > .15$ for all tests. EN-CBM = Early Numeracy Curriculum-Based Measure; SAT-10 = Stanford Achievement Test, 10th Edition.
[a]Items were rated as 0 (*no*), 0.5 (*partially*), or 1 (*yes*).
[b]Overall fidelity of implementation was rated from 1 (*low*) to 7 (*high*).
[c]Quality-of-instruction items were rated from 1 (*not present*) to 4 (*highly present*).

–0.23), and teachers providing individual turns ($\beta = -0.23$) and (b) SAT-10 outcomes and modeling ($\beta = 0.26$). Moderate negative associations were found between gains in (a) EN-CBM outcomes and implementation of Activities 1 through 3, modeling, providing individual turns, and overall fidelity of implementation ($\beta = -0.39$, $\beta = -0.23$, $\beta = -0.23$, and $\beta = -0.20$, respectively) and (b) SAT-10 outcomes and providing group response opportunities and quality of instruction ($\beta = -0.21$ and $\beta = -0.26$, respectively).

Discussion

We examined the impact of a Tier 2 first-grade intervention program targeting critical whole-number content. We hypothesized that at-risk students in the intervention condition would show greater gains than their at-risk peers in the control condition. The results from this study provide partial support for our primary hypothesis. On a proximal measure assessing conceptual understanding of whole-number content, ProFusion, students showed statistically significantly greater gains and a large effect (WWC, 2011). Results on a proximal measure of procedural fluency, EN-CBM, and a distal measure of conceptual understanding, SAT-10, were not statistically significant, but both showed small positive effect sizes. The WWC (2011) provides a classification system to generate an overall descriptor of results when a student has multiple measures. If a study has one statistically significant positive result and the other results in the study are nonsignificant but show positive effect sizes, the overall results for the study are described as having a statistically significant positive impact on student outcomes. Although there are a limited number of randomized controlled trials focused on early mathematics (Dyson, Jordan, & Glutting, 2013), the pattern of results found in this study is similar to results found in other studies of comprehensive intervention programs with first-grade students in which overall positive results were found with greater impacts on proximal measures of achievement (Bryant, Bryant, Gersten, Scammacca, Funk, et al., 2008; Bryant et al., 2011; Fuchs et al., 2005; Rolfhus et al., 2012). For example, Fuchs et al. (2005) found effect sizes up to 0.7 on proximal measures of achievement, but in a replication study of the same program, Rolfhus et al. (2012) found an effect size of 0.34 on a distal measure of achievement.

Regarding our second research hypothesis that higher levels of implementation fidelity would be associated with greater student gains, an analysis of the association between implementation variables and student outcomes did not show significant results. This analysis was designed to serve as a proxy for the potential role of teacher–student interactions functioning as a mediator of student outcomes. The pattern of nonsignificant mixed results that were found across associations makes it difficult to draw conclusions concerning our hypothesis that greater levels of implementation quality would be positively associated with student achievement gains.

Limitations

A number of considerations are important when interpreting the results of this study. First, because the study is an initial pilot study, the sample is restricted by geographic location and the demographic characteristics of the study sample are not representative of the national population of first-grade students. In addition, because of the small sample size, the power to detect treatment impact is limited. Sufficient power may not have been present to detect small positive trends in the data on the EN-CBM and SAT-10 measures. Despite these limitations, there is some preliminary support, on the basis of the overall findings on student outcomes measures, that Fusion had a positive impact on student achievement.

A critical key in supporting our theory of change is whether teacher and student behaviors mediate achievement. However, because implementation data were not collected in control classrooms, we were not able to directly examine mediation. To navigate this barrier, we examined implementation fidelity based on the hypothesis that teachers who implemented Fusion with greater fidelity would have greater rates of critical behaviors (e.g., providing models of math concepts) than teachers who implemented with lower levels of fidelity. Thus, although we present fidelity-of-implementation data as an attempt to explore the role of mediation, it is a limitation of the study that formal mediation was not conducted.

Our exploratory analysis found no significant results supporting our hypothesis that higher levels of implementation quality would mediate student outcomes. In part, this may have been because of overall high levels of implementation across groups, and thus range restriction may have attenuated the magnitude of those relationships. The mean score on each of the five implementation variables examined was at least 0.9 (on a scale of 0–1, with 1 representing full implementation) or higher, and the largest standard deviation was 0.2. In other words, because all teachers implemented Fusion to a relatively high level, there was a lack of variability in measuring implementation quality. The lack of variability may have contributed to the interesting pattern of negative associations between implementation fidelity and the EN-CBM and SAT-10, findings that are counterintuitive. That is, greater implementation was associated with lower outcomes on those measures. In part, this may be because of the small sample size of the study and the fact that a small sample size may have limited the stability of estimates. Another possibility is that there may have been poor alignment between the EN-CBM and SAT-10, measures and the content of the intervention. The EN-CBM measures were designed to be a proximal measure of procedural fluency, but the fluency focus of the Fusion intervention was aimed at fluency with basic facts and not the skills directly assessed by the EN-CBM measures (e.g., number identification and magnitude comparison). The same concern holds for the SAT-10, which was designed as a distal measure of achievement. However, the SAT-10 included content such as geometry and measurement concepts that were not a direct focus of the Fusion intervention. We do caution that because the results were nonsignificant, any interpretation of the results and potential causes is speculative.

Implications for Practice and Future Research

Examining results from this study and other research studies on early mathematics intervention within a framework provided by a theory-of-change model offers insights to guide both research and practice. Given that the results presented were generated from a pilot study, caution should be implied when interpreting results and discussing implications for practice. The results from this study and, in particular, the overall positive impact results on student outcomes offer a starting point for schools to consider the use of Fusion as one potential tool within an RtI service delivery model. Additional evidence is needed to warrant a definitive statement on whether schools should implement Fusion. One important consideration when evaluating the student outcome results from this study is how the results fit within a pattern of curriculum design and research findings in early mathematics. Across an array of early mathematics intervention programs (e.g., Bryant et al., 2011; Dyson, Jordan, & Gluting, 2013; Fuchs et al., 2005, Sood & Jitendra, 2013), there is a general trend by researchers who develop curricula toward building curricula based on the two intervention components that provide the theoretical foundation for Fusion—a focus on whole-number content and an explicit and systematic instruction approach. Given that these approaches mirror recommendations from individual experts (Gersten et al., 2009; Milgram & Wu, 2005) and national panels (NMAP, 2008), it is reasonable to suggest that schools should actively look for intervention programs with a similar focus given that current programs may not meet these recommendations. Schools may need to provide support to educators in the classroom as they implement current intervention programs lacking these features. For example, schools can support the collection of observation data focused on key instructional behaviors and link observation findings to teacher coaching and professional development in critical areas such as providing students opportunities to verbalize their mathematical thinking (Clarke et al., 2008). This is a role well suited for school psychologists or other school professionals with expertise in conducting classroom observations and instructional design. In addition, as new intervention programs are developed and made available, educators looking to adopt programs should ensure that their review process puts particular emphasis on examining whether programs under consideration provide focused content coverage and an explicit and systematic instructional design approach.

Although there is general consensus that curricula should contain the two aforementioned components, future research should take a more fine-grained approach to tease out and manipulate specific program variables. One line of research in this vein examines the concept of instructional intensity (Warren, Fey, & Yoder, 2007) by focusing on specific instructional design manipulations that vary the intensity of the instructional experience for the student. For example, Bryant et al. (2011), extending a line of research working with at-risk first graders (Bryant, Bryant, Gersten, Scammacca, & Chavez, 2008; Bryant, Bryant, Gersten, Scammacca, Funk et al., 2008), have—over multiple iterations of their program—focused on increasing instructional intensity by expanding the amount of instructional time required as part of the intervention. Although increasing instructional intensity can be accomplished by varying delivery parameters such as group size and number of lessons, instructional intensity can also be increased by manipulating elements embedded within intervention programs such as the number of models provided by the teacher or individual response opportunities for students.

To address the issue of examining instructional intensity as a potential mediating variable, one possible remedy would be to use a more robust observational system across treatment and control classrooms. For example, in a recent efficacy trial of a kindergarten mathematics program, our research team conducted approximately 400 observations in 129 kindergarten classrooms using a low-inference observation instrument called the Classroom Observations of Student-Teacher Interactions–Mathematics (COSTI-M; Doabler et al., in press). We were particularly interested in using the COSTI-M to examine the relationship between the rate of explicit instructional interactions and student mathematics achievement (Doabler et al., in press). Specifically, the COSTI-M allowed us to capture three key components of instructional interactions hypothesized to potentially mediate student mathematics achievement: explicit teacher demonstrations, student practice opportunities, and timely academic feedback. A key finding from the study is that students in classrooms with higher rates of practice opportunities made substantively important gains in critical mathematics outcomes. Further research on Fusion and other programs using a similar observation system framework would allow a more robust examination of potential mediation variables and shed light on the theories of change underlying different programs.

Fuchs et al. (2013) have conducted a line of longitudinal research focused on developing and investigating mathematics interventions across the early elementary grades. This line of research offers a number of valuable insights. Although the intervention programs studied were effective as measured by traditional analytic approaches, they were not universally effective for all students in two critical ways. First, although some students responded to the program, the impact on achievement was not great enough to fully reduce the achievement gap between at-risk students and not at-risk peers. This finding mirrors similar results from other studies of curriculum programs (e.g., Clarke et al., 2011) and the general difficulty in fully reducing achievement gaps (Starkey & Klein, 2008). Second, and even more critically, despite the provision of research-based instruction, there remained a subgroup of students who did not respond to the intervention. One potential way to increase the efficacy of instructional interventions for students who do not respond is to modify programs based on a theory-of-change model. For example, to increase the efficacy of a preexisting intervention program, Fuchs et al. (2010) modified a portion of the intervention program focused on increasing procedural fluency with mathematic facts, an area of particular difficulty and critical importance for students with mathematics learning disabilities (Geary, 2004). One set of modifications was linked to a theory of change based on the potential moderating role of what the authors termed *domain general abilities*. Specifically, Fuchs et al. (2013) structured the math fact practice time to "compensate for . . . potential weaknesses in the domain general abilities associated with difficulty with math facts: inattentive behavior, processing speed, phonological processing, working memory, and reasoning ability" (p. 260). In part, these types of systematic manipulations are an inherent and valuable part of design science (Brown, 1992; Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003) and the construction of research-validated curricula (Clements, 2007).

These efforts illustrate that multiple perspectives and theories of changes are informing the development and research of mathematics curricula. It is not to suggest that research teams investigating a specific theory of change and corresponding mediator and moderator variables fail to acknowledge the role of other potential mediators and moderators but, rather, specific research studies and lines of research may focus on delving deeply and systematically into a specific component of a broader theory-of-change model. Thus, as a whole, researchers should be prepared to assimilate findings from these various research lines into more robust theories of change that enhance the overall quality of curriculum development efforts. A number of next steps are vital to extend the research on Fusion. Foremost is the need to collect implementation data and teacher–student interaction data across a condition to allow true mediation analysis. Second, greater attention should be paid to ensuring that the student outcome measures more closely align with the underlying theory of change. This could include a proximal measure of procedural fluency focused on basic number combination fluency and a distal measure of conceptual understanding with an emphasis on whole-number concepts and understanding. Lastly, given that this pilot study had a small sample size, subsequent studies should examine Fusion with larger sample sizes and test Fusion across an array of different geographic locations and sample demographic characteristics to increase the generalizability of results.

Conclusion

Given the critical importance of a successful start in mathematics (Hanich, Jordan, Kaplan, & Dick, 2001) and the need for effective intervention programs for use with tiered models of instruction (Gersten et al., 2009), the importance of researchers developing, investigating, and modifying theory-of-change models is paramount. Through individual and collective efforts to do so, the research field has the opportunity to contribute greatly to the quality of mathematics instruction provided in our nation's schools as we attempt to provide all students with a strong foundation in mathematical understanding.

References

Archer, A. L., & Hughes, C. A. (2010). *Explicit instruction: Effective and efficient teaching.* New York: Guilford Press.

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

Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of Learning Disabilities, 38*, 333–339. doi:10.1177/00222194050380040901

Bransford, J., & Donovan, M. S. (2005). Scientific inquiry and how people learn. In M. S. Donovan & J. Bransford (Eds.), *How students learn: Mathematics, and science in the classroom* (pp. 397–420). Washington, DC: National Academies Press.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of Learning Sciences, 2*, 141–178.

Bryant, D. P., Bryant, B. R., Gersten, R., Scammacca, N., & Chavez, M. (2008). Mathematics intervention for first and second grade students with mathematics difficulties: The effects of tier 2 intervention delivered as booster lessons. *Remedial and Special Education, 29*, 20–32.

Bryant, D. P., Bryant, B. R., Gersten, R. M., Scammacca, N. N., Funk, C., Winter, A., & Pool, C. (2008). The effects of tier 2 intervention on the mathematics performance of first-grade students who are at risk for mathematics difficulties. *Learning Disability Quarterly, 31,* 47–63.

Bryant, D. P., Bryant, B. R., Roberts, G., Vaughn, S., Pfannenstiel, K. H., Porterfield, J., & Gersten, R. (2011). Early numeracy intervention program for first-grade students with mathematics difficulties. *Exceptional Children, 78,* 7–23.

Clarke, B., Baker, S. K., & Chard, D. J. (2008). Best practices in mathematics assessment and intervention with elementary students. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (Vol. V, pp. 453–464). Bethesda, MD: National Association of School Psychologists.

Clarke, B., & Shinn, M. R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review, 33*(2), 234–248.

Clarke, B., Smolkowski, K., Baker, S. K., Fien, H., Doabler, C. T., & Chard, D. J. (2011). The impact of a comprehensive tier 1 core kindergarten program on the achievement of students at-risk in mathematics. *Elementary School Journal, 111,* 561–584.

Clements, D. H. (2007). Curriculum research: Toward a framework for "research-based curricula." *Journal for Research in Mathematics Education, 38,* 35–70.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32,* 9–13.

Common Core State Standards Initiative. (2010). *Common core standards for English language arts & literacy in history/social studies, science, and technical subjects.* Retrieved from Common Core State Standards Initiative: Preparing America's Students for College & Career website: http://www.corestandards.org/assets/CCSSI_ELA%20Standards.pdf

Coyne, M. D., Kame'enui, E. J., & Carnine, D. W. (2011). *Effective teaching strategies that accommodate diverse learners* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Doabler, C., & Nelson-Walker, N. (2009). *Ratings of classroom management and instructional support.* Unpublished observation instrument, Center on Teaching and Learning, College of Education, University of Oregon, Eugene, OR.

Doabler, C. T., Baker, S. K., Kosty, D., Smolkowski, K., Clarke, B., Miller, S. J., & Fien, H. (2013). *Examining the association between explicit mathematics instruction and student mathematics achievement.* Manuscript submitted for publication.

Doabler, C. T., Nelson-Walker, N., Kosty, D., Fien, H., Baker, S. K., Smolkowski, K., & Clarke, B. (in press). Examining teachers' use of evidence-based practices during core mathematics instruction. *Assessment for Effective Intervention.*

Dyson, N. I., Jordan, N. C., & Glutting, J. (2013). A number sense intervention for low-income kindergartners at risk for mathematics difficulties. *Journal of Learning Disabilities, 46*(2), 166–181.

Feingold, A. (2009). Effect sizes for growth-modeling analysis for controlled clinical trials in the same metric as for classical analysis. *Psychological Methods, 14,* 43–53.

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97,* 493–513.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2013). Intervention effects for students with comorbid forms of learning disability: Understanding the needs of nonresponders. *Journal of Learning Disabilities, 46,* 534–548.

Fuchs, L. S., Fuchs, D., & Zumeta, R. O. (2008). Response to intervention: A strategy for the prevention and identification of learning disabilities. In E. L. Grigorenko (Ed.), *Educating individuals with disabilities: IDEIA 2004 and beyond* (pp. 115–135). New York: Springer.

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.

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

Gersten, R., & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *The Journal of Special Education, 33,* 18–28.

Gersten, R. M., Beckmann, S., Clarke, B., Foegen, A., March, L., Star, J. R., & Witzel, B. (2009). *Assisting students struggling with mathematics: Response to intervention (RTI) for elementary and middle schools* (Practice Guide Report No. NCEE 2009–4060). Washington, DC: U.S. Department of Education, National Center for Education Evaluation and Regional Assistance.

Ginsburg, H. P., & Baroody, A. J. (2003). *Test of early mathematics ability* (3rd ed.). Austin, TX: ProEd.

Glover, T. A., & DiPerna, J. C. (2007). Service delivery for response to intervention: Core components and directions for future research. *School Psychology Review, 36,* 526–540.

Good, R. H., Gruba, J., & Kaminski, R. A. (2002). Best practices in using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in an outcomes-driven model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology IV* (pp. 679–700). Washington, DC: National Association of School Psychologists.

Hanich, L. B., Jordan, N. C., Kaplan, D., & Dick, J. (2001). Performance across different areas of mathematical cognition in children with learning difficulties. *Journal of Educational Psychology, 93,* 615–627.

Harcourt Educational Measurement. (2002). *Stanford Achievement Test [SAT-10].* San Antonio, TX: author.

Individuals with Disabilities Education Improvement Act of 2004, P.L. 108–446, 20 U.S.C. § 1400 *et seq* (2004).

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80,* 437–447.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States.* Mahwah, NJ: Lawrence Erlbaum Associates

MacKinnon, D. P., & Luecken, L. J. (2008). How and for whom? Mediation and moderation in health psychology. *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association, 27,* S99–S100.

Milgram, R. J., & Wu, H. S. (2005). *The key topics in a successful math curriculum*. Retrieved June 24, 2013, from http://math.berkeley.edu/~wu/

Morgan, P. L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities, 42*, 306–321.

National Association of State Directors of Special Education. (2006). *Response to intervention: Policy considerations and implementation.* Alexandria, VA: author.

National Center for Education Statistics. (2013). NAEP 2013 mathematics: Report card for the nation and the states. Washington, D. C.: U. S. Department of Education, Institute of Education Sciences.

National Council of Teachers of Mathematics. (2013). *2013 NCTM legislative platform.* Reston, VA: author.

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

National Research Council. (2001). *Adding it up: Helping children learn mathematics.* Washington, DC: Mathematics Learning Study Committee.

National Research Council. (2009). *Mathematics learning in early childhood: Paths towards excellence and equity.* Washington, DC: National Academies Press.

National Science Board. (2008). *Science and engineering indicators 2008* (NSB 08–01 [volume 1] and NSB 08–01A [volume 2]). Arlington, VA: National Science Foundation.

Newman-Gonchar, R., Clarke, B., & Gersten, R. (2009). *A summary of nine key studies: Multi-tier intervention and response to interventions for students struggling in mathematics*. Portsmouth, NH: RMC Research, Center on Instruction.

Obama, B. H. (2013, February 12). Obama's state of the union. *The New York Times.* Retrieved from http://www.nytimes.com

Okamoto, Y., & Case, R. (1996). Exploring the microstructure of children's central conceptual structures in the domain of number. *Monographs of the Society for Research in Child Development, 61*, 27–59. doi: 10.1111/j.1540-5834.1996.tb00536.x

Rolfhus, E., Gersten, R., Clarke, B., Decker, L., Wilkins, C., & Dimino, J. (2012). *An evaluation of Number Rockets: A Tier 2 intervention for grade 1 students at risk for difficulties in mathematics* (NCEE 2012–4007). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Rothman, A. J. (2013). Exploring connections between moderators and mediators: Commentary on subgroup analyses in intervention research. *Prevention Science, 14*, 189–192.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–177.

Shinn, M. R. (1989). *Curriculum-based measurement: Assessing special children.* New York: The Guilford Press.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* Thousand Oaks, CA: Sage.

Sood, S., & Jitendra, A. (2013). An exploratory study of number sense program to develop kindergarten students number proficiency. *Journal of Learning Disabilities, 46,* 328–346.

Starkey, P., & Klein, A. (2008). Sociocultural influences on young children's mathematical knowledge. In Saracho, O. N., & Spodek, B (Eds.), *Contemporary perspectives on mathematics in early childhood education* (pp. 253–276). Charlotte, NC: Information Age.

Vaughn, S., & Fuchs, L. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research & Practice, 18*, 137–146.

Vaughn, S., Wanzek, J., Woodruff, A. L., & Linan-Thompson, S. (2007). Prevention and early identification of students with reading disabilities. In D. Haager, J. Klinger, & S. Vaughn (Eds.), *Evidence-based reading practices for response to intervention* (pp. 11–27). Baltimore, MD: Paul H Brookes.

Wanzek, J., & Vaughn, S. (2010). Research-based implications from extensive early reading interventions. In T. A. Glover & S. Vaughn (Eds.), *The promise of response to intervention: Evaluating current science and practice* (pp. 113–142). New York: Guilford Press.

Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research Reviews, 13*, 70–77.

What Works Clearinghouse. (2011). *What Works Clearinghouse: Procedures and standards handbook (version 2.1).* Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Witzel, B. S., Riccomini, P. J., & Schneider, E. (2008). Implementing CRA with secondary students with learning disabilities in mathematics. *Intervention in School and Clinic, 43*, 270–276.

Woodcock, R. M., & Johnson, M. B. (1989). Woodcock-Johnson Psycho-Educational Battery–Revised. Allen, TX: DLM Teaching Resources.

Wu, H. (1999). Basic skills versus conceptual understanding. *American Educator, 23*(3), 14–19.

Date Received: July 8, 2013
Date Accepted: February 3, 2014
Guest Editor: Robert Volpe is cited. ■

Ben Clarke, PhD, is a Research Assistant Professor with the Center on Teaching and Learning at the University of Oregon. He has served as a Principal Investigator on nine federally funded research grants in the area of mathematics. Dr. Clarke is interested in the development of mathematical thinking and instructional methods and assessment approaches to support student acquisition of mathematics content.

Christian T. Doabler, PhD, is a Research Assistant Professor with the Center on Teaching and Learning at the University of Oregon. Currently, he serves as a Co-Principal Investigator on three IES-funded grants focused on early mathematics instruction and intervention. He specializes in curriculum design, classroom observation systems, and the prevention of learning difficulties.

Mari Strand Cary, PhD, received her doctorate in Child Psychology from the University of Minnesota in 2004. She is currently a research associate at the Center on Teaching and Learning at the University of Oregon. Much of her work focuses on utilizing educational technology, particularly in the areas of math and science, to deliver instructionally-sound, motivating, and individualized learning experiences to elementary students.

Derek Kosty is a doctoral candidate at the University of Oregon and holds a position as a data analyst at the Oregon Research Institute. He has experience in quantitative analysis of experimental and quasi-experimental trials in education and psychology. His research interests include measurement of effective instructional practices in early mathematics and reading.

Scott Baker, PhD, is the executive director of the Center on Research and Evaluation (CORE) at Southern Methodist University (SMU). Dr. Baker is interested in the impact of interventions on child outcomes, the mechanisms that underlie effective interventions, and how intervention impact varies by factors intrinsic and extrinsic to the child.

Hank Fien, PhD, is a Research Associate Professor and Co-Director of the Center on Teaching and Learning at the University of Oregon. Currently, he serves as principal Investigator or co-Principal Investigator on ten federally-sponsored research and development grants studying the feasibility and efficacy of early reading and early math curricula and interventions. His research interests include reading and mathematics development in young children, integrating instructional design and gaming technology, and empirically validating interventions aimed at preventing or ameliorating student academic problems.

Keith Smolkowski, PhD, is a research scientist at Oregon Research Institute. His research focuses on the children's social and academic behavior from early childhood through adolescence and the practices, interventions, and supports that lead to improvements in student achievement, social-emotional skills, and behavior. He received his doctorate in Special Education from the University of Oregon.