

Manipulating Treatment Dose: Evaluating the Frequency of a Small Group Intervention Targeting Whole Number Operations

Robin S. Coddling

University of Minnesota

Amanda M. VanDerHeyden

Education & Research Consulting, Inc.

Ryan J. Martin, Sheila Desai, Noelle Allard, and Leigh Perrault

University of Massachusetts Boston

Treatment dose is an understudied aspect of treatment effectiveness. This study compared the frequency with which a small-group mathematics intervention was delivered weekly (i.e., four times, twice, once) with a control condition while controlling for total duration. 101 at-risk students in grades 2–4 were randomly assigned to a condition following universal screening and skill-based assessments. Multilevel modeling was used to evaluate final score and growth on three measures. Results suggested that for the most proximal computation measure, treatment sessions occurring four times weekly produced clear benefits. On the application measure, students in all treatment groups outperformed students in the control condition. For the most complex computation measure, frequency was not a useful predictor. Grade was a moderating variable.

A central tenet of multitiered systems of service delivery, such as Response to Intervention (RtI), is the notion that the intensity of resources expended, time committed, and frequency of data collected all increase as students proceed from less-intensive to more-intensive levels of support (Batsche et al., 2005). It has been suggested that the technical adequacy of these layered models of prevention and intervention depends on effectively adjusting and aligning intervention intensity according to student needs (Barnett, Daly III, Jones, & Lentz, 2004). Although intervention intensity is a multifaceted construct, a consistent aspect of the definition is time allocated to intervention supports (Mellard, McKnight, & Jordan, 2010). It is commonly assumed that more overall time dedicated to intervention supports reflects higher levels of intensity, and that such intensity results in better outcomes for children with more risk for academic failure (e.g., Gersten et al., 2009). Unfortunately, evidence suggests that schools do not differentiate time allocated to interventions across tiers, likely because time is a fixed commodity (Mellard et al., 2010).

One of the most common barriers to implementation of RtI reported by teachers is the time, scheduling, and staff resources required to provide intervention supports (Castro-Villareal, Rodriguez, & Moore, 2014; Swanson, Solis, Ciallo, & McKenna, 2012). Schools may find it particularly challenging to designate time and resources to the delivery

of tier 2 intervention supports, which are intended to be short-term, targeted interventions that address specific student skills in small groups (Johnson, Carter, & Pool, 2012; Mellard & Johnson, 2008). In fact, intervention integrity is a common problem in schools, with teachers reporting intervention time and duration as a critical barrier to implementation (Long et al., 2016). In particular, scheduling time to allocate to the delivery of intervention services is reported by teachers as challenging (Castro-Villareal et al., 2014; Swanson et al., 2012). Implicit within evidence-based practice is that intervention procedures are not only effective but also efficient, suggesting that identifying intervention strategies that address students' needs within the time restrictions that schools face is paramount (VanDerHeyden & Harvey, 2012). Only a handful of studies have empirically examined the RtI framework in mathematics, and even fewer studies have investigated tier 2 intervention practices (Johnson, Carter, & Pool, 2012; Newman-Gonchar, Clarke, & Gersten, 2009). Consequently, research is needed to examine the manner in which treatment time, or dose, can be allocated when delivering tier 2 intervention supports in mathematics.

Treatment Dose

Two of the most commonly described aspects of intervention intensity in schools are the frequency and duration of treatment sessions (Coddling & Lane, 2015). Determining

Requests for reprints should be sent to Robin S. Coddling, University of Minnesota. Electronic inquiries should be sent to rcoddling@umn.edu

the amount of treatment necessary to produce meaningful changes in the performance of students who are at-risk for poor mathematics outcomes might offer a more efficient and precise mechanism for arranging service delivery. Treatment dose is often conceptualized as: (a) number of learning trials per session, (b) minutes of instruction per session, (c) frequency of instructional sessions, and (d) total duration over which the intervention was delivered (Coddling & Lane, 2015). Treatment intensity can be varied along any of these dimensions and potentially adjusted. For example, a small number of minutes per session can be allocated for more days per week over a shorter duration, or a greater number of minutes per session can be allocated for fewer weekly sessions over a longer duration. Systematic examination of aspects of treatment dose that take into consideration educational recommendations for effective learning opportunities as well as evidence from the distributed practice literature may be useful to inform the scheduling of intervention delivery in schools.

Plenty of evidence exists recommending the use of frequent and brief learning opportunities to promote skill acquisition and fluency across academic areas, including mathematics (e.g., Coddling, Burns, & Lukito, 2011; Martens et al., 2007). Learning opportunities are ideally sequenced systematically according to students' instructional level and adjusted as students master each skill set (Fuchs et al., 2008; Hasselbring, Goin, & Bransford, 1988; Martens & Eckert, 2007). Conducting a survey level assessment using curriculum-based assessment (Gickling & Havertape, 1981) is one way that interventionists can identify students' instructional levels on a predetermined skill hierarchy (Burns, Coddling, Boice, & Lukito, 2010). Intervention sessions are constructed with both skills in isolation (e.g., flash cards) as well as application of component skills in context (e.g., word problems) (Daly, Martens, Barnett, Witt, & Olson, 2007; Fuchs et al., 2008). Finally, sessions are structured to incorporate modeling, immediate feedback on performance, and reinforcement for effort (Coddling et al., 2011; Daly et al., 2007; Fuchs et al., 2008).

Within the cognitive literature, it is well established that distributed practice outperforms massed practice by almost half a standard deviation, illustrating the benefits of such practice on learning and memory (Donovan & Radosevich, 1999). Distributed practice refers to sessions or trials that are separated by intervals of time (Cepeda, Pashlwer, Vul, & Wixted, 2006). These spaced sessions have been shown to improve the retention of learned material among college students (Rohrer & Taylor, 2006, 2007). Practice that is not separated by time intervals and instead occurs during a single uninterrupted session is described as massed practice. However, there are nuances in the distributed practice literature, such that brief intervals between sessions result in greater retention for simple tasks than for more complex tasks, for which longer time lags between sessions may be more useful (Donovan & Radosevich, 1999). Given that time is scarce in schools and constitutes one of the most critical barriers impacting effective intervention implementation, modifying the frequency of the intervention, as opposed to the overall duration, in accordance with the theory of distributed practice might be useful.

Treatment Frequency and Mathematics Interventions

We are only aware of three studies that have systematically examined treatment frequency as applied to mathematics interventions. Duhon, Mesmer, Atkins, Greguson, and Olinger (2009) investigated whether increasing the frequency of daily sessions beyond what was already provided during classroom instruction would result in improved performance. Using a multiple baseline design across three students at-risk for mathematics difficulties, Duhon et al. (2009) demonstrated that additional time allocated to brief opportunities for daily practice produced improvement in computation fluency for all students and was maintained for two students. In a follow-up study, Schutte et al. (2015) randomly assigned 48 third graders to one of three conditions that varied the frequency with which basic addition facts were practiced: (a) one massed session of four 1-min trials, (b) two sessions consisting of two 1-min trials practiced twice daily, and (c) four 1-min sessions conducted four times daily. Significant differences in final performance and session growth were observed favoring both distributed practice groups as compared to the massed practice group. These findings suggest that one reasonable option for addressing initial non-response to mathematics intervention is to increase the frequency of treatment sessions provided. However, the conclusions that can be drawn from these studies are limited by the focus on a singular mathematics skill addressed through one instructional strategy.

Using a standardized treatment package, Topping et al. (2011) compared the impact of intensity on the outcomes of a randomized control trial examining a class-wide peer tutoring program implemented either once or three times weekly for 30 min per session. No significant differences were detected, suggesting that the frequency of sessions was not important. This outcome could be specific to the program itself, suggesting that exposure produced positive outcomes regardless of the frequency of weekly sessions. In contrast to the research by Duhon et al. (2009), Topping et al. (2011) did not include a daily treatment option. Further interpretation is limited by the lack of control for total weekly treatment duration (Topping et al., 2011). Holding total time constant within each week is important to isolate the effect of intervention frequency from that of total treatment exposure.

These preliminary studies suggest that increasing the frequency of an intervention might be an appropriate solution for students that are not responding to the core curriculum when number combination fluency is the outcome. However, treatment duration was not held constant in the study by Topping and colleagues (2011), and none of the studies represented tier 2 intervention delivery with small groups of students. Although recommendations exist suggesting that tier 2 services be delivered four times weekly (Gersten et al., 2009), the frequency of these sessions has not been systematically examined to our knowledge. It is also unclear whether differing levels of task complexity are affected by treatment frequency. The cognitive literature suggests that the type of task and the length of intervals between instructional sessions are moderating variables (Donovan & Radosevich, 1999). The benefit of distributed practice is particularly robust for

simple tasks when brief intertrial intervals are planned. However, for more complex tasks these findings are less consistent, and longer intertrial breaks may be more effective.

Purpose

The purpose of the current study was to extend the literature evaluating the impact of treatment frequency on the outcomes of mathematics interventions by providing an intervention package that addressed whole number operations to students who were at-risk for mathematics failure. The same intervention protocol was delivered in small groups of three to five students each (Gersten et al., 2009) across the three treatment conditions: (a) four sessions weekly, (b) two sessions weekly, and (c) one session weekly. Treatment conditions were compared to a no-treatment control. The total minutes of weekly treatment exposure were held constant (i.e., 48 min).

Our primary research question examined whether differences in mathematics outcomes would be observed among the four variations of treatment frequency across three different outcome measures that varied in task complexity: fact families, application skills, and grade-level complex computation problems. Given the evidence indicating that brief, frequent opportunities for practice lead to greater benefits than massed practice for simple tasks (Donovan & Radosevich, 1999; Martens et al., 2007), we anticipated that students in the four-times-weekly condition would outperform students in the other conditions on the fact family measures. However, for the remaining two outcome variables it was anticipated that all three treatment groups would outperform the control group, but differences among the treatment conditions may not be as substantial given the greater complexity of the tasks.

METHOD

Participants

A total of 236 students from a small urban elementary school in the northeastern region of the United States were recruited for possible participation. The IRB granted passive consent since all students in grades 2, 3, and 4 were screened for participation. Letters were sent home to all parents, and parents who did not want their child to participate could opt out. No parents opted out of the study. The initial recruitment sample consisted of 85 2nd-grade students in four general education classrooms, 80 3rd-grade students in three classrooms, and 71 4th-grade students in three classrooms. The following data apply to the entire recruitment sample because the school did not have students' individual demographic information accessible for final participants. Fifty-one percent of students in the school were male. Forty-eight percent of students were from low-income families, of which 41 percent qualified for free lunch. For 35 percent of the students, English was not the first language, and 15 percent of the students were English-language learners. The racial composition of the students was 53 percent Caucasian, 20 percent African American, 16 percent Hispanic, 7 percent Asian, and 4 percent Multiracial.

Only students that met the inclusion criteria were included in the study. From this school, 141 (61 percent) students in grades 2–4 were considered at-risk for mathematics difficulties, as indicated by grade specific Curriculum-Based Measurement-Mathematics (M-CBM; VanDerHeyden, 2014) data, and were eligible for participation in the study. Fluency criteria from Deno and Mirkin (1977) were applied whereby students performing in the frustration and instructional levels were considered at-risk (<40 digits correct for grades 2 and 3; <80 digits correct for grade 4). Following a survey-level assessment of isolated computation skills, 40 students were eliminated due to (a) very low scores on the survey-level assessment (i.e., frustration range of performance on all grade-level skills), (b) failure to correctly represent a computation problem using drawings (line representations; Ginsburg, 2009), or (c) participation in special education services (Shapiro, 2010). The final sample consisted of 101 students from ten different classrooms across grades. Between six and 16 students participated from each classroom. Participants included 39 2nd-graders, 46 3rd-graders, and 16 4th-graders.

Setting

Intervention sessions took place within spare classrooms, the cafeteria (nonmeal times), and vacant office spaces within the school. *Everyday Mathematics*, 3rd Edition (McGraw Hill Education, 2012), served as the core mathematics curriculum. Students received 45 min of core instruction (Tier 1) daily. On the statewide mathematics accountability test conducted the year prior to data collection, 46 percent of 3rd-grade students (4th graders by commencement of the study) scored below the proficient range. Data were not available for other students because accountability testing commences in third grade. The school did not engage in universal screening or apply a multitiered model to identify and intervene with students struggling with mathematics.

Interventionists and Training

Screening and progress monitoring data were collected by eight school psychology graduate students (3 doctoral, 5 specialist-level). Six of these students (3 doctoral, 3 specialist-level) also led the intervention groups for each condition. Each graduate student was trained to use curriculum-based assessment and implement academic interventions as part of their program coursework. Graduate students participated in two training sessions that were specific to the procedures employed in the current study. First, graduate students viewed a 20-min video, in which the first author modeled the treatment protocol with three graduate students (not otherwise serving in the study) acting as participants. Graduate students observed the video in pairs and practiced implementing the treatment protocol with one another. During the second session, the first author reviewed all screening materials and observed test administration to correct errors and provide feedback. Scripted protocols guided all assessment and intervention activities.

Measures

Two forms of curriculum-based assessment tools were used: (a) subskill mastery measures and (b) general outcome measures. Subskill mastery measures, referred to as curriculum-based measurement mathematics (M-CBM), were used for instructional planning and identification of students' appropriate instructional level (Hosp & Ardoin, 2008). These measures contained one or two specific skills, permitting a precise evaluation of students' proficiency with several basic and complex computation skills. General outcome measures sampled items across year-end grade level expectations. Two tools were used: one that addressed grade level application skills (i.e., Monitoring Basic Skills Progress-Basic Concepts and Applications), and another that addressed computation skills (i.e., Monitoring Basic Skills Progress- Basic Math Computation).

Curriculum-Based Measurement-Mathematics (M-CBM)

Computation probes consisted of two worksheets containing 30–35 total problems and two problem types. The 2nd grade probe sampled addition and subtraction fact families with numerals 0–20. Third grade probes contained mixed 2- and 3-digit addition and subtraction problems with and without regrouping. Fourth-grade probes consisted of multiplication and division fact families. Probes were obtained using an online generator and represent critical indicators of computation skills (VanDerHeyden, 2014). Previous research yielded adequate psychometric evidence for these M-CBM tools (Burns, VanDerHeyden, & Jiban, 2006; VanDerHeyden & Burns, 2008). Delayed alternate-form reliability for the probes is .85, and the probes correlated with scores on the Stanford Achievement Test (Harcourt, 1997) in the low to moderate range ($r = .27$ to $.40$), typical of similar tools (Foegen, Jiban, & Deno, 2007). Scores consisted of digits computed correctly per 2-min administration.

Monitoring Basic Skills Progress-Basic Concepts and Applications (MBSP-APP)

MBSP-APP (Fuchs, Hamlett, & Fuchs, 1999) is a curriculum-based measure designed to assess applied mathematics skills for grades 1–6. Probes for grades 2–4 include grade specific problems pertaining to counting, number concepts, number names/vocabulary, measurement, charts and graphs, money, fractions, applied computation, and word problems. Standardized administration procedures and scoring procedures were followed. MBSP-APP demonstrates adequate reliability and validity. Internal consistency coefficients exceed .94 for grades 2, 3, and 4. Validity coefficients with CBT/McGraw Hill (1997) total math scores range from .74 to .81 across grades 2 through 4. Concurrent validity coefficients with MBSP-Computation range from .63 to .81 across grades 2 through 4. Scores consisted of number of correct digits in the answer.

Monitoring Basic Skills Progress-Basic Math Computation (MBSP-COMP)

MBSP-COMP (Fuchs, Hamlett, & Fuchs, 1999) is a curriculum-based measure designed to assess mathematics computation skills according to annual curriculum standards for grades 1–6. Specific skills assessed by MBSP-COMP vary according to grade level and include simple and complex (multidigit) computation problems using whole and rational numbers. There were 25 problems on each probe, and two probes were administered to ensure students would not complete the task before time elapsed. For 2nd-graders, probes contained simple and complex (3-digit by 1-digit, 2-digit by 2-digit, and 3-digit by 2-digit) addition and subtraction whole number problems. For 3rd-graders, probes contained complex addition and subtraction as well as simple multiplication and division whole number problems. For 4th graders, probes contained simple and complex multiplication and division whole number problems as well as complex addition and subtraction whole number problems and simple rational number problems. Standardized directions for administration and scoring were followed and specified answer keys used. The MBSP-COMP probes demonstrate adequate reliability and validity. Alternate-form reliability coefficients for students without disabilities range from .73 to .81 across grades 2–4. Reliability coefficients for students without disabilities range from .81 to .86. Validity coefficients are .82 with Math Computation Test scores (Fuchs, Fuchs, Hamlett, & Stecker, 1991), .66 with the numbers and concepts subtest scores on the Stanford Achievement Test (Gardner, Rudman, Karlsen, & Merwin, 1982), and .67 with math computation subtest scores of the Stanford Achievement Test (Gardner et al., 1982). Scores consisted of points awarded for each correct answer blank, with the exception of word problems and money problems for which one point was awarded per digit in the answer.

Survey-Level Assessment

Using procedures recommended by Shapiro (2010) and criteria developed by Deno and Mirkin (1977), a survey-level assessment using subskill mastery curriculum-based measures printed from an online generator (VanDerHeyden, 2014) was used to determine students' appropriate instructional level. Assessments used to further define students' skill strengths and weaknesses are useful for promoting treatment validity (Fuchs, Fuchs, & Speece, 2002; Hosp & Ardoin, 2008). Subskill mastery measures are useful for instructional intervention planning because these measures represent a subdivision of broad grade-level skills into smaller slices of the curriculum (Hosp & Ardoin, 2008; Shapiro, 2010). A survey level assessment is a process whereby performance on any one skill is analyzed to determine skill proficiency, monitor progress, and guide instruction (Hosp, Hosp, & Howell, 2007). The scores (recorded as digits correct per 2 min) obtained from administration of the subskill mastery measures can then be compared to criteria to determine whether performance falls within the frustration, instructional, or mastery level of skill performance (Burns et al., 2006; Deno & Mirkin, 1977;

TABLE 1
Hierarchical Skill Sequence

<i>Operations and Skills</i>	<i>Grade 2</i>	<i>Grade 3</i>	<i>Grade 4</i>
Addition			
Sums to 6	X	X	
Sums to 12	X	X	
Sums to 20	X	X	X
2×2 Digit with and without regrouping		X	
Subtraction			
Minuends to 5	X	X	
Minuends to 9	X	X	
Minuends to 20	X	X	X
2×2 Digit with and without regrouping		X	
Addition + Subtraction			
Fact Families			X
Multiplication			
Factors to 5			X
Factors to 12			X
Division			
Divisors to 5			X
Divisors to 12			X

Howell & Nolet, 2000). The lowest skill in the sequence that is not mastered (either falling within the frustration or instructional range) was the first target skill addressed by the intervention (Shapiro, 2010). For students in grades 2 and 3, mastery criteria were 40 or more digits correct per 2 min (DC), instructional level was 20 to 39 DC, and frustration level was 19 or fewer DC. For fourth graders, mastery was 80 DC, the instructional level ranged between 40 and 79 DC, and the frustration level was 39 or fewer DC. For these assessment probes, a sequence of skills was developed for each grade using Common Core State Standards in Mathematics (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010; VanDerHeyden, McLaughlin, Algina, & Snyder, 2012).

Table 1 illustrates the operations and skill hierarchy that was administered to students in each grade during the survey-level assessment. For second grade, the survey-level assessment contained addition and subtraction facts. For third grade, the survey-level assessment consisted of basic addition and subtraction facts as well as double-digit addition and subtraction problems with and without regrouping. The survey-level assessment for 4th-grade students contained addition and subtraction fact families as well as basic multiplication and division facts.

To rule out conceptual misunderstandings, students were asked to represent a number combination visually by using line drawings (Ginsburg, 2009). This task consisted of asking students to draw a solution to the first equation on the skill identified as the appropriate instructional-level. Drawings were used so that students could illustrate their mathematical thinking and reasoning. Students with substantial difficulty with underlying mathematical reasoning were not appropriate candidates for the intervention.

Materials

Each student was provided with a folder containing intervention materials as well as a pencil, a red pen, a dry-erase

marker, and a dry-erase board. The interventionists each had a binder with a schedule, intervention protocols, record forms, flashcards, a stopwatch, an audio recorder, tokens, and prizes.

Experimental Conditions

Treatment Frequency

Students in all treatment conditions were randomly assigned to homogenous groups of 3 to 5 students according to results of the survey-level assessment. (The randomization process is described below in the procedures section.) Across groups, practice with application tasks was grade-specific and content remained constant; however, a mastery model was applied to computation practice such that the survey-level assessment data were used to identify the first skill in a sequence that would be targeted for intervention (Hasselbring et al., 1988; Shapiro, 2010). Once each group of 3–5 students reached a median score of 40 DC (for grades 2 and 3) or 80 DC (for grade 4) on a 2-min targeted mastery probe (see intervention steps below), then a new skill was introduced during the subsequent day on which the intervention was delivered.

The treatment protocol was 12 min in length and was shown to be effective in previous research (VanDerHeyden et al., 2012). First, students participated in guided practice of mathematics facts with the interventionist or in dyads for 3 min on a scripted rotating schedule (e.g., choral responding, white boards, peer practice). Mathematics facts were presented to the students on flash cards; students engaged in choral response by calling out the answers simultaneously or writing their answers on individual dry-erase boards and holding them in the air. Students were also divided into dyads to practice with flashcards. Students were instructed to take turns as the “player” (i.e., the student that is responding) and the “coach” (i.e., the student that presents the math fact). The coach provided corrective feedback if the player did not know the answer, created a pile of known and unknown facts, and reviewed any unknown facts with the player. The interventionist instructed students to switch roles halfway through the guided practice portion of the treatment protocol. For complex computation problems, guided practice was conducted with worksheets.

Second, students engaged in 2 min of independent practice. Students were provided with mathematics worksheets featuring the targeted problems (i.e., as determined by the survey-level assessment) and were encouraged to beat their scores from the previous session in order to receive a token that would later be exchanged for a small reward. At the end of 2 min, students were allotted 3 min to correct and score their own worksheets. The interventionist read the answers aloud to the group and the students marked problems on their worksheets as correct or incorrect. If a problem was incorrect, students crossed out the incorrect number(s) and replaced it with the correct number(s). The interventionist verified and recorded scores and distributed tokens to students who scored higher than in the previous session.

Third, interventionists facilitated guided practice on applied mathematics problems for 4 min. Applied problems were derived from alternative forms of the MBSP-APP

measure according to each grade level (e.g., word problems, measurement, charts and graphs). Interventionists modeled and briefly reviewed a problem type to the group and then provided students with the opportunity to work on a similar problem independently. The interventionist provided direct instruction (e.g., breaking down the problem into smaller steps, demonstrating and modeling the strategy to solve the problem) and immediate, corrective feedback (i.e., ensuring students applied the appropriate strategies; correcting errors in the process of problem solving or with the outcome) as needed. Students were encouraged to help one another if the interventionist was providing individualized support to another student. When the 4 min expired, students exchanged any earned tokens for prizes before returning to their classrooms.

Control

Students assigned to the control condition did not receive additional treatment beyond standard classroom instruction with the *Everyday Mathematics* curriculum. These students participated in grade-wide screening conducted before commencement of the study and weekly progress monitoring but were not provided with any formal math intervention services.

Procedure

Screening occurred during two consecutive Fridays two weeks prior to the start of the intervention. All students in grades 2, 3, and 4 were assessed with grade-specific M-CBM, MBSP-APP, and MBSP-COMP probes. All students participated in a survey-level assessment to determine final eligibility for the intervention, homogenous student groups (i.e., students from the same grade with the same starting skill were grouped together), and a starting skill for the intervention.

Participants were assigned to one of four conditions (i.e., four times weekly, twice weekly, once weekly, control) using stratified random sampling, controlling for skill level and grade so that each condition consisted of approximately equivalent numbers of students from each grade and skill level. Six groups of 3–5 students were arranged per condition (i.e., 24 total groups). Each condition consisted of (a) three groups of 3rd-graders (for two groups, the starting skill was double-digit addition, and for one group, the starting skill was addition facts with sums to 20), (b) two groups of 2nd-graders (starting skills were sums to 12 and sums to 20, respectively), and (c) one group of 4th-graders whose starting skill was basic multiplication facts. Students in all conditions (except the control) received a total of 192 min of intervention over 4 weeks. In the first condition ($n = 26$), students received one 12-min treatment session (the length of one complete trial of the treatment protocol) four days per week (Monday through Thursday). In the second condition ($n = 26$), students received two consecutive treatment sessions (trials of the treatment protocol), lasting a total of 24 min twice per week (Tuesday and Thursday). In the third condition ($n = 23$), students received four consecutive

treatment sessions lasting 48 min once per week (Thursday). The fourth condition ($n = 26$) served as the control group. Two interventionists were assigned to each of the three treatment conditions ($n = 6$). Each interventionist was responsible for running three groups within the same treatment condition.

Class-wide progress monitoring was conducted once per week by graduate students. Progress monitoring occurred each Friday morning throughout the study (i.e., four consecutive weeks). Graduate students visited each classroom and administered all three outcome measures (i.e., M-CBM, MBSP-APP, MBSP-COMP).

Procedural Fidelity

Direct Observation of Student Engagement

Student engagement was measured using the Behavioral Observation of Students in Schools (BOSS; Shapiro, 2004). The first author observed 4 interventionists for 3 sessions each, and 2 interventionists were observed twice, representing 12 percent to 18 percent of total intervention sessions. A round-robin method was applied, and the observer recorded students' on-task behavior during 15-s intervals using momentary time sampling. Mean observed student engagement (active + passive engagement) across interventionists was 98 percent (*range*, 94 percent to 100 percent), with students actively participating in 87 percent (*range*, 72 percent to 94 percent) of intervals observed.

Observation of Treatment Protocol Steps

Procedural fidelity of intervention steps was evaluated using a checklist that accounted for 18 distinct interventionist behaviors representing one 12-min treatment session. All intervention sessions for each group were recorded using handheld audiotape recorders. Two specialist-level school psychology graduate students were trained in the intervention procedures and acted as independent reviewers. These graduate students were not otherwise involved with the study. The observers listened to 54 audio recordings (i.e., 12-min sessions) representing between 2 and 6 days of treatment for each of the six interventionists (57 percent of treatment sessions administered).

Mean treatment adherence was 93 percent (*range*, 60 percent to 100 percent) as measured across all six interventionists. There were no commission or omission errors. Errors reflected wording alterations when intervention activities were explained. Mean treatment adherence was (a) 100 percent and 86 percent (*range*, 66 percent to 100 percent) for the four-times-weekly interventionists, (b) 97 percent (*range*, 89 percent to 100 percent) and 87 percent (*range*, 60 percent to 100 percent) for the twice-weekly interventionists, and (c) 100 percent, respectively, for both once-weekly interventionists.

Interscorer Agreement

Interscorer agreement was computed for 20 percent of the total progress monitoring sessions by an independent scorer.

TABLE 2
Pre- and Posttest Descriptive Statistics According to Measure and Experimental Condition

Experimental Condition	M-CBM	MBSP-APP	MBSP-COMP
Pretest			
Four Times Weekly	7.88 (5.96)	11.16 (5.07)	5.92 (3.68)
Twice Weekly	8.16 (5.53)	11.24 (5.75)	5.12 (4.82)
Once Weekly	7.68 (5.02)	12.59 (6.58)	5.45 (3.73)
Control	7.88 (5.88)	10.96 (5.30)	4.36 (3.15)
Posttest			
Four Times Weekly	24.44 (15.10)	23.20 (7.82)	15.33 (11.54)
Twice Weekly	17.08 (14.04)	23.29 (9.54)	11.76 (9.38)
Once Weekly	16.57 (13.96)	25.20 (9.42)	12.10 (10.92)
Control	16.76 (13.87)	20.26 (8.49)	14.04 (10.50)

Note. Standard deviations are listed parenthetically. M-CBM = Curriculum-based measurement in mathematics. MBSP-APP = Monitoring Basic Skills Progress-Basic Concepts and Applications. MBSP-COMP = Monitoring Basic Skills Progress-Basic Math Computation.

The independent scorer was a school psychology graduate student with course experience in M-CBM who also assisted with procedural fidelity calculations. Mean interscorer agreement for M-CBM and MBSP-COMP was computed on a digit-by-digit basis and was 97 percent (range, 60 percent to 100 percent) and 96 percent (range, 57 percent to 100 percent), respectively. Discrepancies with identifying student handwritten numbers were a common error. For MBSP-APP, mean agreement was 90 percent (range, 62 percent to 100 percent). Failure to give partial credit for word problems was a common error.

RESULTS

First, descriptive statistics are provided for the pretest performance on the outcome variables and the mean number of skills mastered during intervention. Second, results from the multilevel modeling analyses are provided. Three different multilevel models were conducted representing each measure.

Descriptive Statistics

Prior to implementing study procedures, a one-way between-groups MANOVA was performed to investigate group differences across the outcome measures: (a) M-CBM, (b) MBSP-COMP, and (c) MBSP-APP. No statistically significant differences were found between groups across the dependent measures, $F(9, 279) = 0.42, p = .92$. Table 2 presents the means and standard deviations according to experimental condition.

The mean number of skills mastered across the course of the intervention for the four-times-weekly and twice-weekly groups was 2 (range, 1–3) and for the once-weekly group was 1.67 (range, 1–2). Figure 1 represents the percentage of students in the risk range on each of the three measures. Risk status was established using criteria by Deno and Mirkin

(1977) for M-CBM and according to the MBSP manual (i.e., less than or equal to the 25th percentile; Fuchs et al., 1999). The percentage of students considered to be at-risk declined across each outcome measure, regardless of group assignment. Across all three measures, students in the four-times-weekly treatment group experienced the largest percentage decrease relative to other conditions.

Missing data for M-CBM ranged from 4 percent to 8 percent ($Mdn = 4$ percent), with 85 percent of students having complete data across all five weeks. For MBSP-APP, missing data ranged from 4 percent to 12 percent ($Mdn = 6$ percent), and 82 percent of students had complete data. For MBSP-COMP, missing data ranged from 4 percent to 13 percent ($Mdn = 6$ percent), and 82 percent of students had complete data. All data were missing at random and were due to student absences.

Multilevel Modeling

Multilevel modeling was used to examine predictors of slope and final score across all measures. SAS PROC MIXED was used for the analyses due to the longitudinal nature of the data and our interest in monitoring individual progress of each student. Two-level models were analyzed where students served as the level-2 unit and repeated observations (1 pretest screening + 4 weeks of consecutive progress monitoring) served as the level-1 unit. Across all models, the restricted maximum likelihood (REML) estimation method and unstructured covariance were employed. Predictors for the models were grade and treatment assignment (i.e., four times weekly, twice weekly, once weekly, or control). Treatment assignment was coded (1 to 4) so that variables representing group membership could be used to predict differences in students' final scores and growth rates over time. The control group and grade 4 students were used as the reference groups. For parsimony, the 1st-level equation of students' progress over time was combined with the 2nd-level equation of differences in students' performance as a function of a treatment assignment and grade (Singer & Willett, 2003). Progress monitoring scores were centered at the end of treatment (final session at end of five weeks), resulting in negative numbers when growth was positive.

Table 3 displays the unconditional and final models across measures. The unconditional model was fit in order to determine whether students' performances varied over time across each measure. The random effects parameters demonstrated that across measures, students' scores significantly varied around the mean, and there were significant differences between each student's observed and predicted scores over time. Final models were determined to be a better fit than the unconditional models because AIC and BIC values were lower (Singer & Willett, 2003).

For M-CBM, the time parameter showed that, on average, students were gaining 2.73 DC from session to session (the negative number reflects centering at the end of treatment). The parameter for final status indicated that the mean score at the end of treatment across grades and conditions was 23.52 DC. ICC calculations indicated that 64 percent of score variation occurred across students (Peugh, 2010).

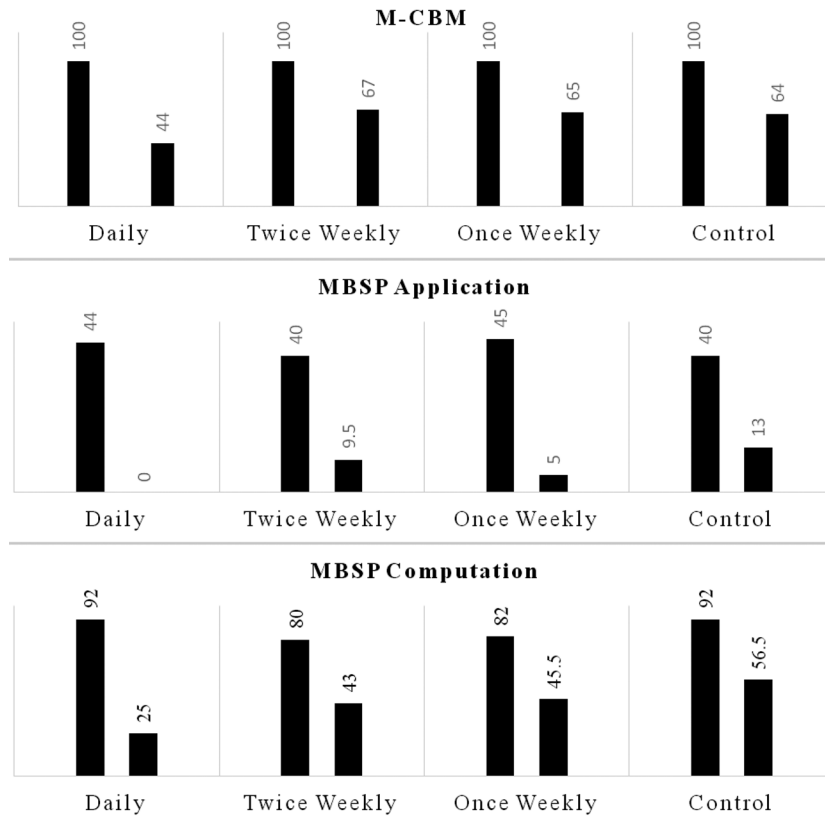


FIGURE 1 Each panel represents the percentage of participants scoring in the at-risk range during pre (week 1) and post (week 5) test. The first panel represents curriculum-based measures in mathematics (M-CBM) results, the second panel represents Monitoring Basic Skills Progress application (MBSP-APP) and the third panel represents Monitoring Basic Skills Progress computation (MBSP-COMP).

In the final model, grade and treatment condition were significant predictors of the intercept and slope. Participation in the four-times-weekly treatment condition was a significant predictor of final score and growth. Follow-up tests indicated that students in the four-times-weekly treatment condition performed significantly better than those in the control ($t = 3.61, p = 0.0004$), twice-weekly ($t = 2.49, p = 0.0131$), and once-weekly ($t = 3.59, p = 0.0004$) conditions. Fourth graders yielded significantly higher final scores than second and third graders, and third graders had the lowest final scores. With respect to slope, the four-times-weekly group yielded higher rates of growth than students in the control conditions. Fourth graders displayed significantly more growth than second and third graders. Level 1 (residual) and Level 2 (intercept) variance estimates decreased 17 percent and 60 percent, respectively, after adding grade and condition to the model (Peugh, 2010).

For MBSP-APP, students on average were gaining 2.82 points from session to session. The average final score across grades and conditions was 26.41. ICC calculations indicated that 68 percent of score variation occurred across students (Peugh, 2010). In the final model, both grade and treatment condition were significant predictors of the intercept and slope. Participation in any treatment condition (i.e., four times, twice, or once weekly)

significantly predicted final score. Follow-up comparison tests yielded no significant differences among the treatment conditions. Second grade was a significant predictor of final score, with students performing lower than fourth graders. Only the twice-weekly condition was a significant predictor of growth. Fourth graders displayed significantly more growth than second and third graders. Variance estimates decreased 0.03 percent and 26 percent, respectively, after adding grade and treatment (Peugh, 2010).

For MBSP-COMP, students, on average, were gaining 1.81 points from session to session. The average final score across grades and conditions was 13.65. ICC calculations indicated that 67 percent of score variation occurred across students (Peugh, 2010). In the final model, only grade was a significant predictor for the intercept and slope. Fourth graders displayed higher final scores and more growth than second and third graders. Students in the once- and twice-weekly conditions displayed final scores that were over 1.5 and nearly 2 points lower than controls, whereas students in the four-times-weekly condition yielded final scores that were 1.55 points higher than controls. Follow-up tests revealed that students in the four-times-weekly group significantly outperformed students in the once- ($t = 2.03, p = 0.04$) and twice-weekly conditions ($t = 2.30, p = 0.02$) but not the control ($t = 1.03, p = 0.30$). Variance estimates

TABLE 3
Multilevel Modeling

	<i>M-CBM</i>		<i>MBSP-APP</i>		<i>MBSP-COMP</i>	
	<i>Unconditional Model</i>	<i>Final Model</i>	<i>Unconditional Model</i>	<i>Final Model</i>	<i>Unconditional Model</i>	<i>Final Model</i>
Fixed Effects						
Final Score	23.52 (1.09)***	43.73 (2.44)***	26.41 (0.75)***	29.44 (1.99)***	13.65 (0.78)***	32.25 (1.60)***
Time	-2.73 (0.21)***	-5.47 (0.55)***	-2.82 (0.14)***	-3.30 (0.39)***	-1.81 (0.16)***	-4.96 (0.41)***
Condition						
Four Times		8.28 (2.30)**		3.90 (1.88)*		1.55 (1.51)
Twice Weekly		2.54 (2.31)		3.82 (1.90)*		-1.93 (1.53)
Once Weekly		-0.22 (2.37)		4.74 (1.96)*		-1.63 (1.57)
Grade						
2		-20.00 (2.47)***		-11.50 (2.01)***		-20.73 (1.62)***
3		-33.50 (2.42)***		-3.67 (1.97)		-22.47 (1.58)***
Time X Condition						
Four Times		-1.86 (0.52)**		-0.67 (0.37)		-0.007 (0.39)
Twice Weekly		-0.66 (0.53)		-0.78 (0.37)*		-0.73 (0.39)
Once Weekly		-0.14 (0.54)		-0.64 (0.39)		-0.78 (0.41)
Time x Grade						
2		3.27 (0.56)***		1.46 (0.39)**		3.26 (0.42)***
3		4.81 (0.55)***		0.99 (0.38)*		3.43 (0.41)***
Random Effects						
Intercept	73.88 (11.72)***	29.45 (5.33)***	35.67 (5.56)***	26.11 (4.31)***	46.10 (7.21)***	17.31 (3.11)***
Residual	41.31 (3.01)***	33.88 (2.48)***	17.08 (1.25)***	16.49 (1.22)***	22.63 (1.66)***	18.69 (1.38)***
Model Summary						
AIC	3369.50	3188.7	2920.2	2858.7	3037.7	2861.3
BIC	3374.70	3193.9	2925.4	2863.9	3043.0	2866.5
Deviance Statistics	3365.50	3184.7	2916.2	2854.7	3033.7	2857.3
Estimated Parameters	4	12	4	12	4	12

Note. Parameter estimate standard errors listed in parentheses.

*** $p < .0001$.

** $p < .01$.

* $p < .05$.

decreased 17 percent and 62 percent, respectively, after adding grade and treatment condition (Peugh, 2010).

DISCUSSION

The purpose of this study was to evaluate the differential impact of treatment session frequency on three outcomes when the treatment itself, total number of treatment weeks, and number of weekly intervention minutes were held constant. Our hypothesis that the four-times-weekly treatment condition would produce the highest final student scores and the most growth over time for the simplest outcome measure, M-CBM, was confirmed. Results for the distal and more complex outcome measures were mixed. For MBSP-APP, participation in any of the treatment groups produced higher final scores than the control. The twice-weekly group exhibited more growth than the control group, with little difference between the other conditions. However, given the lack of differences between the four-times-weekly and twice-weekly group, support seems to favor the delivery of the briefer, more frequent treatment sessions when considering the time constraints that schools and teachers face. For MBSP-COMP, session frequency was not a significant predictor. Grade was a moderator of outcomes.

Treatment Frequency

The clearest support for shorter, more frequent intervention sessions versus longer, weekly sessions was displayed on the M-CBM outcome. This measure was most closely aligned with the intervention scope and sequence and represented the most basic mathematics task (i.e., focusing on combinations of key number operations). This finding reflects cognitive learning theory indicating that distributed practice outperforms massed practice, particularly for simple tasks (Donovan & Radosevich, 1999). These findings also support previous literature suggesting that brief practice opportunities with modeling, feedback, and reinforcement lead to gains in computational fluency (Bryant et al., 2011; Coddling et al., 2011; VanDerHeyden et al., 2012). Finally, these data provide support for recent recommendations that intervention sessions be delivered at least four times weekly to at-risk students (Gersten et al., 2009).

The impact of exposure to the intervention, regardless of intensity, on the application measure is likely because the same types of problems on this measure were explicitly practiced as part of the intervention, and we are unaware whether the control group was exposed to these problem types in the general curriculum. Therefore, for the MBSP-Application measure, something, (i.e., the intervention

provided four times weekly, twice weekly, or weekly) was indeed better than nothing (i.e., control). This outcome could also be explained by the fact that 61 percent of all students in grades 2 through 4 were considered to be at-risk according to the screening measures we employed, suggesting that changes in core instruction might be warranted. For example, traditional mathematics instruction provides limited opportunities for directly teaching word problem-solving (Woodward et al., 2012). Our outcomes also indicated that the twice-weekly group exhibited significantly higher session gains than the control group. This finding might be supported by the literature suggesting that longer intertrial breaks for more complex tasks result in better outcomes when practice is distributed (Donovan & Radosevich, 1999).

MBSP-COMP consisted of a variety of problems that students may or may not have been exposed to as part of the intervention, making it the most distal and complex outcome measure. When compared to the control group, there were no significant differences in final score or growth for any of the treatment conditions. Interestingly, students in the four-times-weekly condition significantly outperformed students in the once- and twice-weekly groups. It is possible that the lack of differences between treatment and control groups was the result of the small sample size. Another explanation for the failure to find differences compared to controls may be that students achieved mastery of only an average of 1.5 (i.e., once-weekly condition) to 2 subskills (i.e., twice- & four-times-weekly conditions). It might not be expected that gains will be observed on this more distal outcome measure until mastery over a greater number of subskills is attained. There is some evidence suggesting that fluency gains with simple computation problems improves complex procedural calculations (e.g., Coddling, Chan-Iannetta, Palmer, & Lukito, 2009; Fuchs, Fuchs, & Compton, 2012), but that may not occur until students have achieved a sufficient level of subskill fluency (VanDerHeyden & Burns, 2009).

This lack of transfer may also suggest that the strength of the intervention could be enhanced. There are several recommended aspects of mathematics interventions that were not included in the intervention protocol (Gersten et al., 2009). These include (a) introducing a small, focused set of number combinations each session (Harniss, Stein, & Carnine, 2002), (b) teaching the same big conceptual idea across fluency and application activities (Harniss et al., 2002), and (c) incorporating cumulative review (Mayfield & Chase, 2002). It is also possible that use of research assistants, as opposed to teachers or teaching assistants, may have resulted in poorer outcomes. There is mixed evidence suggesting that teachers, with more pedagogical and content knowledge and skill, provide more comprehensive feedback to students (Campbell & Malkus, 2011; Desimone, Smith, & Phillips, 2013; Gersten et al., 2009; National Mathematics Advisory Panel, 2008).

Grade

These findings need to be considered in context with students' grade levels. Across both computation measures, fourth graders outperformed second and third graders and also displayed more growth over time. Higher final scores

on M-CBM for fourth graders might simply reflect higher expected rates of fluency (Burns et al., 2006). That is, the mastery level for fourth grade is over 80 correct digits, compared to 40 digits correct for second and third graders (Deno & Mirkin, 1977). It also possible that the intervention was more appropriately matched with fourth-grade instructional content, or that those students had greater conceptual understanding of number operations than younger students. Third graders persistently yielded the lowest final scores and the fewest gains per week across the two computation measures. A potentially important difference in the subskill sequence for third graders exposed to this intervention was the presence of multidigit computation. Most students whose entry point for the intervention was multidigit addition only mastered this one skill during the course of treatment. This finding may suggest that (a) the intervention was not adequate to address conceptual principles and procedural skills necessary for solving multidigit computation problems, (b) underlying prerequisite skills were not well established, or (c) multidigit computation is a more challenging skill requiring more time for students to master than single-digit number operations.

With respect to the application measure, fourth graders performed significantly higher at the end of treatment than second but not third graders. Fourth graders also displayed significantly higher rates of growth over the course of treatment. Although these measures were grade-specific, the high level of risk identified in this sample overall suggests that perhaps students did not have a deep enough conceptual understanding of basic mathematical principles to generalize these ideas to the application problems at the second grade level.

Limitations and Future Directions

There are several limitations that could inform future research on treatment intensity of mathematics interventions. First, the intervention was only conducted over four weeks, and the common conceptualization of intervention length within an RtI structure is 6 to 12 weeks (e.g., Gersten et al., 2009). It is possible that differences yielded between conditions across outcome measures may have changed if the intervention was provided for a greater number of weeks. Second, participants were sampled from a school that exhibited high risk for mathematics failure according to our screening measures as well as statewide achievement test data. It is possible that the outcomes of our small-group intervention were impacted by the school's need to make adjustments to core instruction. Third, this school was not employing an RtI framework. Different outcomes may be achieved with samples generated from schools employing multitiered models. Fourth, as already noted, research assistants rather than teachers or other school personnel conducted intervention sessions. The use of school personnel may have resulted in different outcomes given their prior experience with the students and familiarity with the core curriculum. Additionally, each research assistant implemented only one condition. However, we included two measures of procedural fidelity (student engagement and procedural adherence) which suggested that there were not meaningful differences

across interventionists. Fifth, the acceptability of this intervention program or the treatment schedule was not analyzed, and it is unknown whether it would be feasible for local school personnel to employ. Sixth, the sample size was relatively small; therefore, future research might consider replicating findings with a larger sample.

Conclusion

The results of this study yield preliminary support for the use of brief, frequent treatment sessions to improve whole-number operations for students determined to be at-risk for mathematics difficulties. In our study, this was particularly the case for the simplest outcomes most aligned with treatment. An essential decision within multitiered systems of support is the distribution of finite intervention resources. This study demonstrated the value of using the same number of supplemental instructional minutes, but changing the distribution to maximize student learning outcomes. We also verified support for a brief 12-min intervention protocol that employs recommended active treatment ingredients (Coddling et al., 2011; Fuchs et al., 2008; VanDerHeyden et al., 2012). Schools seeking to prevent and repair mathematics achievement deficits should consider briefer, more frequent intervention sessions, particularly to improve basic computation skills, a key area of need for students at-risk for mathematics failure (Gersten & Chard, 1999; Stickney, Sharp, & Kenyon, 2012).

REFERENCES

- Barnett, D. W., Daly III, E. J., Jones, K. M., & Lentz Jr., F. (2004). Response to intervention: empirically based special service decisions from single-case designs of increasing and decreasing intensity. *Journal of Special Education, 38*, 66–79.
- Batsche, G., Elliott, J., Graden, J. L., Grimes, J., Kovalski, J. F., Prasse, D., et al. (2005). *Response to intervention: Policy considerations and implementation*. Alexandria, VA: NASDSE.
- Bryant, D. P., Bryant, B. R., Roberts, G., Vaughn, S., Pfannenstiel, K. H., Porterfield, J., et al. (2011). Early numeracy intervention programs for first-grade students with mathematics difficulties. *Exceptional Children, 78*, 7–23.
- Burns, M. K., Coddling, R. S., Boice, C., & Lukito, G. (2010). Meta-analysis of acquisition and fluency math interventions with instruction and frustration level skills: Evidence for a skill-by-treatment interaction. *School Psychology Review, 39*, 69–83.
- Burns, M. K., VanDerHeyden, A. M., & Jiban, C. (2006). Assessing the instructional level for mathematics: A comparison of methods. *School Psychology Review, 35*, 401–418.
- Campbell, P. F., & Malkus, N. N. (2011). The impact of elementary mathematics coaches on student achievement. *The Elementary School Journal, 111*, 430–454. doi:10.1086/657654
- Castro-Villareal, F., Rodriguez, B. J., & Moore, S. (2014). Teachers' perceptions and attitudes about response to intervention (RTI) in their schools: A qualitative analysis. *Teaching and Teacher Education, 40*, 104–112. doi:10.1016/j.tate.2014.02.004
- Cepeda, N. J., Pashler, H., Vul, E., & Wixted, J. T. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*, 354–380. doi: 10.1037/00333-2909.132.354
- Christ, T. J., Scullin, S., Tolbize, A., & Jiban, C. L. (2008). Implications of recent research: Curriculum based measurement of math computation. *Assessment for Effective Intervention, 33*, 198–205. Common Core State Standards Initiative. (2010).
- Coddling, R. S., Burns, M. K., & Lukito, G. (2011). Meta-analysis of mathematical basic-fact fluency interventions: A component analysis. *Learning Disabilities Research & Practice, 26*, 36–47. doi:10.1111/j.1540-5826.2010.00323.x
- Coddling, R. S., Chan-Iannetta, L., Palmer, M., & Lukito, G. (2009). Examining a class-wide application of cover-copy-compare with and without goal setting to enhance mathematics fluency. *School Psychology Quarterly, 24*, 173–185. doi:10.1037/a0017192
- Coddling, R. S., & Lane, K. L. (2015). A spotlight on treatment intensity: An important and often overlooked component of intervention inquiry. *Journal of Behavioral Education, 24*, 1–10. doi:10.1007/s10864-014-9210-z
- Common Core State Standards Initiative. (2010). Common core state standards for mathematics. Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf
- Daly, E. J., III, Martens, B. K., Barnett, D., Witt, J. C., & Olson, S. C. (2007). Varying intervention delivery in response to intervention: Confronting and resolving challenges with measurement, instruction, and intensity. *School Psychology Review, 36*, 562–581.
- Deno, S. L., & Mirkin, P. K. (1977). *Data-based program modification: A manual*. Reston, VA: Council for Exceptional Children.
- Desimone, L., Smith, T. M., & Phillips, K. R. (2013). Linking student achievement growth to professional development participation and changes in instruction: A longitudinal study of elementary students and teachers in Title I schools. *Teachers College Record, 115*, 1–14.
- Donovan, J. J. & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology, 84*, 795–805.
- Duhon, G. J., Mesmer, E. M., Atkins, M. E., Greguson, L. A., & Olinger, E. S. (2009). Quantifying intervention intensity: A systematic approach to evaluating student response to increasing intervention frequency. *Journal of Behavioral Education, 18*, 101–118. doi:10.1007/s10864-009-9086-5
- Foegen, A., Jiban, C., & Deno, S. (2007). Progress monitoring in mathematics: A review of the literature. *The Journal of Special Education, 41*, 121–139. doi:10.1177/00224669070410020101
- Fuchs, L. S., Fuchs, D., Craddock, C., Hollenbeck, K. N., Hamlett, C. L., & Schatschneider, C. (2008). Effects of small-group tutoring with and without validated classroom instruction on at-risk students' math problem solving: Are two tiers of prevention better than one? *Journal of Educational Psychology, 100*, 491–509. doi:10.1037/0022-0663.100.3.491
- Fuchs, L. S., Fuchs, D., & Compton, D. L. (2012). The early prevention of mathematics difficulty: Its power and limitations. *Journal of Learning Disabilities, 45*, 257–269. doi:10.1177/0022219412442167
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1991). *Math computation test*. Nashville, TN: Vanderbilt University.
- Fuchs, L. S., Fuchs, D., Powell, S. R., Seethaler, P. M., Cirino, P. T., & Fletcher, J. M. (2008). Intensive intervention for students with mathematics disabilities: Seven principles for effective practice. *Learning Disability Quarterly, 31*, 79–92.
- Fuchs, L. S., Fuchs, D., & Speece, D. L. (2002). Treatment validity as a unifying construct for identifying learning disabilities. *Learning Disabilities Quarterly, 25*, 33–45.
- Fuchs, L. S., Hamlett, C. L., & Fuchs, D. (1999). *Monitoring basic skills progress basic math manual*. Austin, TX: Pro-Ed.
- Gardner, E. F., Rudman, H. C., Karlsen, B., & Merwin, J. C. (1982). *Stanford achievement test*. Stanford, CA: Psychological Corp.
- Gersten, R., Beckmann, S., Clarke, B., Fogen, A., Marsh, L., Star, J. R., et al. (2009). *Assisting students struggling with mathematics: Response to intervention (RtI) for elementary and middle schools* (NCEE 2009-4060). Washington, D.C.: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Gersten, R., & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *Journal of Special Education, 33*, 18–28.
- Gickling, E. E., & Havertape, S. (1981). *Curriculum-based assessment (CBA)*. Minneapolis, MN: School Psychology Inservice Training Network.
- Ginsburg, H. P. (2009). The challenge of formative assessment in mathematics education: Children's minds, teachers' minds. *Human Development, 52*, 109–128.
- Harcourt. (1997). *Stanford Achievement Test Series*. San Antonio, TX: Harcourt Brace & Co.

- Harniss, M. K., Stein, M., & Carnine, D. (2002). Promoting mathematics achievement. In M. R. Shinn, H. M. Walker & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 571–587). Washington, D.C.: National Association of School Psychologists.
- Hasselbring, T. S., Goin, L. I., & Bransford, J. D. (1988). Developing math automaticity in learning handicapped children: The role of computerized drill and practice. *Focus on Exceptional Children*, 20, 1–7.
- Hosp, J. L. & Ardoin, S. P. (2008). Assessment for instructional planning. *Assessment for Effective Intervention*, 33, 69–77. doi:10.1177/1534508407311428
- Hosp, M., Hosp, J., & Howell, K. (2007). *The ABCs of CBM: A practical guide to curriculum-based measurement*. New York: Guilford Press.
- Howell, K. W., & Nolet, V. (2000). *Curriculum-based evaluation: Teaching and decision making* (3rd Ed.). Belmont, CA: Wadsworth, Cengage Learning.
- Johnson, E. S., Carter, D. R., & Pool, J. L. (2012). Introduction to the special issue: The critical role of a strong tier 2 system. *Intervention in School and Clinic*, 48, 195–197. doi: 10.1177/1053451212462877
- Long, A. C. J., Hagermoser Sanetti, L. M., Collier-Meek, M., Gallucci, J., Altschaeff, M., & Kratochwill, T. R. (2016). An exploratory investigation of teachers' intervention planning and perceived implementation barriers. *Journal of School Psychology*, 55, 1–26.
- Martens, B., & Eckert, T. (2007). The Instructional Hierarchy as a Model of Stimulus Control over Student and Teacher Behavior: We're Close but are We Close Enough? *Journal of Behavioral Education*, 16(1), 82–90. doi:10.1007/s10864-006-9026-6
- Martens, B. K., Eckert, T. L., Begeny, J. C., Lewandowski, L. J., DiGennaro, F. D., Montarello, S. A., et al. (2007). Effects of a fluency-building program on the reading performance of low-achieving second and third grade students. *Journal of Behavioral Education*, 16, 39–54. doi:10.1007/s10864-006-9022-x
- Mayfield, K. H., & Chase, P. N. (2002). The effects of cumulative practice on mathematics problem solving. *Journal of Applied Behavior Analysis*, 35, 105–123. doi:10.1901/jaba.2002.35-105
- McGraw Hill Education. (2012). *Everyday Mathematics* (3rd Ed.). New York: McGraw Hill.
- Mellard, D., McKnight, M., & Jordan, J. (2010). RTI tier structures and instructional intensity. *Learning Disabilities Practice*, 25, 217–225.
- Mellard, D. F., & Johnson, E. S. (2008). *RTI: A practitioner's guide to implementation*. Thousand Oaks, CA: Corwin Press.
- National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). *Common Core State Standards Mathematics*. Washington, DC: National Governors Association Center for Best Practices, Council of Chief State School Officers.
- National Mathematics Advisory Panel. (2008) *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education. Retrieved from <http://www2.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf>.
- 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 Corporation, Center on Instruction.
- Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology*, 48, 85–112. doi:10.1016/j.jsp.2009.09.002
- Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practice on the retention of mathematics knowledge. *Applied Cognitive Psychology*, 20, 1209–1224. doi:10.1002/acp.1266
- Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, 35, 481–498. Doi:10.1007/s11251-007-9015-8
- Schutte, G. M., Duhon, G. J., Solomon, B. G., Ponce, B. C., Moore, K., & Story, B. (2015). A comparative analysis of massed vs. distributed practice on basic math fact fluency growth rates. *Journal of School Psychology*, 53, 149–159. doi:10.1016/j.jsp.2014.12.003
- Shapiro, E. S. (2004). *Academic skills problems: Third edition workbook*. New York: Guilford.
- Shapiro, E. S. (2010). *Academic skills problems: Direct assessment and intervention* (4th Edition). New York: Guilford.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.
- Stickney, E. M., Sharp, L. B., & Kenyon, A. S. (2012). Technology-enhanced assessment of math fact automaticity: Patterns of performance for low- and typically achieving students. *Assessment for Effective Intervention*, 37, 84–94. doi:10.1177/1534508411430321
- Swanson, E., Solis, M., Ciullo, S., & McKenna, J. W. (2012). Special education teachers' perceptions and instructional practices in response to intervention implementation. *Learning Disabilities Quarterly*, 35, 115–126.
- Topping, K. J., Miller, D., Murray, P., Henderson, S., Fortuna, C., & Conlin, N. (2011). Outcomes in a randomized controlled trial of mathematics tutoring. *Educational Research*, 53, 51–63. doi:10.1080/00131881.2011.552239
- VanDerHeyden, A. M., (2014). InterventionadvisorSM. Retrieved from <http://interventionadvisor.com>.
- VanDerHeyden, A. M., & Burns, M. K. (2008). Examination of the utility of various measures of mathematics proficiency. *Assessment for Effective Intervention*, 33, 215–224
- VanDerHeyden, A. M., & Burns, M. K. (2009). Performance indicators in math: Implications for brief experimental analysis of academic performance. *Journal of Behavioral Education*, 18, 71–91. doi:10.1007/s10864-009-9081-x
- VanDerHeyden, A. M. & Harvey, M. (2012). Using data to advance learning outcomes in schools. *Journal of Positive Behavior Interventions*, 15, 205–213.
- VanDerHeyden, A. M., McLaughlin, T., Algina, J., & Snyder, P. (2012). Randomized evaluation of a supplemental grade-wide mathematics intervention. *American Education Research Journal*, 49, 1251–1284.
- Woodward, J., Beckmann, S., Driscoll, M., Franke, M., Herzig, P., Jitendra, A., et al. (2012). *Improving mathematical problem solving in grades 4 through 8: A practice guide* (NCEE 2012-4055). Washington, D.C.: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications_reviews.aspx#pubsearch/

About the Authors

Robin S. Coddling, Ph.D., is an Associate Professor of School Psychology at the University of Minnesota. She earned her Ph.D. in school psychology from Syracuse University. Dr. Coddling's research interests focus on the intersection of intervention and implementation by developing and exploring the effectiveness of school-based interventions, the factors that contribute to student responsiveness of those interventions, and strategies to support intervention implementation. Dr. Coddling's work has emphasized academic interventions and associated assessment for data-based decision making, particularly in the area of mathematics.

Amanda M. VanDerHeyden, Ph.D., is a frequent contributor to the literature in the use of data-based decision making to improve efficiency, accuracy, and intensity of instruction and raise schoolwide achievement. She is co-author of the Evidence-Based Mathematics Innovation Configuration for the National Comprehensive Center for Teacher Quality at Vanderbilt University and now the Collaboration for Effective Education Development, Accountability, and Reform at University of Florida. Her most recent effort has been completing a web-based mathematics intervention system that provides screening, progress monitoring,

and multitiered intervention aligned with student need covering numeracy to algebra. This system (Intervention Advisor) is published by TIES and is available September 1, 2016.

Ryan J. Martin, Ph.D., is a postdoctoral fellow and behavioral consultant with May Institute and the National Autism Center, where he currently studies school-based interventions for children with Autism Spectrum Disorder. He earned his Ph.D. in School Psychology from the University of Massachusetts-Boston. Ryan's broader research interests include mathematics intervention, home and school consultation, and methods of assessing and improving treatment fidelity.

Sheila Desai, Ph.D., is a Postdoctoral Fellow in the EASTCONN Psychological and Behavioral Consultation department in Connecticut where she supports students and staff in clinical day treatment settings. She completed her Ph.D. in School Psychology at the University of Massachusetts - Boston. Her primary research interests include training and supervision practices in school psychology.

Noelle Allard is a practicing school psychologist. She earned her specialist degree in School Psychology from the University of Massachusetts-Boston.

Leigh Perrault is a practicing school psychologist. She earned her specialist degree in School Psychology from the University of Massachusetts-Boston.