CLEAR | Center for Leadership Equity and Research

Evaluating The Efficacy of Short-Cycle Mathematic Interventions in a High-Minority/High-Poverty Urban Public Middle School

R. Aaron Wisman, *California State University, Bakersfield*
Marco A. Muñoz, *Jefferson County Public Schools*
P. Tyler Stevens, *Jefferson County Public Schools*

Abstract: There is a national interest among educators about how to properly design, implement, and evaluate academic interventions, particularly as related to students placed at risk. This study provides school-based intervention teams with a rigorous methodological approach for monitoring and evaluating procedural fidelity as well as the effectiveness of short-cycle academic interventions; at the same time, the study describes some of the challenges faced when implementing a tiered academic intervention system. The study adds to the literature by offering empirical evidence for the efficacy of specific instructional strategies aligned to specific elementary and middle-school Common Core Mathematics Standards. A quasi-experimental research design with control groups and pre- and post-tests was used to draw causal inferences about the efficacy of intervention treatments. Two short-cycle mathematics interventions for distinct Common Core Mathematics Standards were found to positively impact student learning. Further implications for practice and future research are discussed.

Keywords: *At risk students, Common Core Mathematics Standards, middle school, quasi-experimental design, Response to Intervention (RtI)*

In the current high-stakes testing accountability environment, education practitioners, in particular those serving schools with high concentrations of students living in poverty and of minority status, are under tremendous pressure to close the achievement gap, especially in reading and mathematics (Every Student Succeeds Act [ESSA] of 2015). The purpose of our study is to (a) investigate the efficacy of a short-cycle tiered mathematics intervention program implemented in a high-minority/high-poverty all-boys urban public middle school in a large Southeastern school district and (b) provide practitioners with a model for monitoring the efficacy and fidelity of any targeted academic intervention. The Response to Intervention (RtI), or Multi-Tiered Systems of Supports (MTSS) model has increased in prevalence across schools and school districts since its inception (Bramlett, Cates, Savina, & Laudinger, 2010; Fuchs & Fuchs, 2006; Hill, King, Lemons, & Partanen, 2012; Klinger & Edwards, 2006; Zirkel & Thomas, 2010).

Riccomini and Witzel (2010) frame their first principle of RtI models as a belief that all students can learn when effective instruction and monitoring are present, providing an axiological stance to guide practitioners in the closure of the achievement gap. Yet many researchers have

critiqued the lack of a sufficient number of studies to guide practitioners in the implementation of interventions, even though practitioners and researchers agree on the paramount importance of intervention research (Strein, Cramer, & Lawser, 2003). For example, Bliss, Skinner, Hautau, and Carroll (2008) demonstrated a decrease in studies of experimental interventions with general education students between 2000 and 2005. Furthermore, Lembke, Hampton, & Beyers (2012) noted the prevalence of reading interventions over those in mathematics. In spite of these and other limitations in the research literature on academic interventions, Dennis (2015) explicitly asserted the promise of RtI as a systematic approach to promote mathematical competence for all students.

Despite the growing interest in intervention studies, many authors note a research-to-practice gap (Bliss, Skinner, Hautau, & Carroll, 2008; Bramlett et al., 2010; Burns, Klingbeil, Ysseldyke, & Petersen-Brown, 2012; Forman, Smallwood, & Nagle, 2005; Landrum & Tankersley, 2004). One reason for this gap, as Bliss et al. (2008) note, may be due to a lack of peer-reviewed "journal articles that directly evaluate, provide evidence for, or empirically validate interventions" (p. 483) and, furthermore, that "researchers often disagree on the appropriate and/or most appropriate process, procedures, and criterion for empirically validating interventions" (p. 484). Other studies note an increase in the prevalence of intervention research, yet make explicit the lack of an increase in the level of rigor of intervention studies (Burns et al., 2012). Another reason for the research-to-practice gap, according to Bramlett et al. (2010), may be due to time constraints preventing practitioners from finding peer-reviewed journal articles every time a problem arises or basic lack of access to peer-reviewed journal articles. As a result, many interventions in practice may not be grounded in research, but rely more heavily on personal experience (Bramlett, Murphy, Johnson, Wallingsford, & Hall, 2002; Burns & Ysseldyke, 2009).

The importance of rigorous intervention evaluation among education practitioners is of paramount importance when allocating precious resources to high-stakes interventions. The present researchers seek to provide a methodological framework for practitioners to ensure an appropriate level of methodological rigor in evaluating the interventions they choose to implement, whether they are research-based or not. We conceptualize methodological rigor using the criteria established by the What Works Clearinghouse (WWC, 2008) with particular attention given to equating intervention and control groups, attrition in the sample, teacher-intervention confound and intervention contamination. The present study also adds to the literature by providing a set of intervention strategies targeting two specific Common Core Mathematics Standards across two grade levels and data upon which to base arguments of generalizability for future study contexts. The following research question is examined: What is the efficacy of a tiered mathematics intervention program implemented in a high-minority/high-poverty, all-boys, urban public middle school in the Southeastern U.S.? The following provides a review of some salient literature to frame the context and need for the present study, an articulation of the methods used to address the research question and results. Implications for practice and future research are discussed.

Conceptualizing a Tiered Mathematics Program

There is evidence that the literature on interventions is growing, especially in light of policies such as the No Child Left Behind Act of 2001 (NCLB), ESSA, the reauthorization of the Individuals with Disabilities Education Act in 2004 (IDEA), and the establishment of the What Works Clearinghouse (WWC) in 2002. However, it seems that researchers have not necessarily increased the level of rigor in research methodologies required by these policies (Burns et al., 2012). The body of literature on interventions is vast and includes many distinct lines of inquiry

with fundamental distinctions such as those between academic and behavioral interventions (Fletcher &Vaughn, 2009), the various tiers of intervention and their interactions (Dennis, 2015; Fuchs, Fuchs, Craddock, Hollenbeck, & Hamlett, 2008; Hill et al., 2012), differences in the ways students are identified for inclusion in higher tiers (Fletcher & Vaughn, 2009), and the methods for evaluating interventions (Barnett et al., 2014; Bliss et al., 2008; Bramlett et al., 2010; Burns et al., 2012) in addition to the myriad intervention practices designed to target specific academic or behavioral outcomes. These hierarchically larger distinctions often contain their own distinct lines of inquiry. For example, among academic interventions, there seems to be more attention given to reading interventions over mathematics interventions. This disparity in the research is not surprising given that the study of academic interventions is rooted in reading (Lembke et al., 2012). There are also differences in the methodological approaches to evaluating interventions or the conceptualization of various intervention practices. For example, Burns et al. (2012) describe different research designs: randomized controlled trials, quasi-experimental designs, and single-case designs. Bramlett et al. (2010) distinguish between specific intervention strategies, such as peer-assisted learning strategies (PALS), pre-teaching, computer assisted instruction, modeling with error correction, intervention intensity, etc.

A Conceptual Framework for Response to Intervention

Credit for the origin of RtI is often given to Deno's (1985) data-based program modification model, or to Bergan's (1977) and Bergan and Kratochwill's (1990) behavioral consultation model. While many variations in the RtI framework have been described (Hoover & Love, 2011), there are some commonalities across RtI models utilized in research and practice. Riccomini and Witzel (2010) describe six principles of RtI emerging from the literature on reading interventions, which are transferable to mathematics intervention models (Lembke, 2012): The belief that all students can learn through the implementation of effective instruction and monitoring of student progress, ensuring all students are periodically screened for increasingly-intensive and individualized tiered supports, using a system of evaluation to monitor the effects of interventions, selecting research-based instructional best practices, differentiating tiers of support structured and staffed by trained educators, and that ongoing program evaluation is implemented across a school or school district. A standard model for an RtI framework elucidates a common structure for targeting individual student needs systematically through a triage approach (Vaughn, Wanzek, Woodruff, & Linan-Thompson, 2006). All students are provided high-quality, research-based tier 1 instruction in the general education setting. Prior to and throughout tier 1 instruction, student progress is monitored and certain individuals are identified for more intensive supports based on evaluation data gleaned from universal screeners. Students identified for more intense and individualized supports are then provided tier 2 instruction, which is again monitored for efficacy; and on into higher tiers of increasing intensity (time, or dosage) and individualization (smaller groups).

Advantages and Challenges to Implementing RtI in Practice and Research

Children encounter many barriers to learning as they progress through their education, including student-input variables, such as socioeconomic background (Sirin, 2005) and process variables, such as the quality of instruction they receive in educational settings (Gersten et al., 2009). Other authors, such as Davis, Herzog, and Legters (2013) describe some typical practices and related challenges to implementing student interventions. Many policies have been enacted that explicitly support school districts in utilizing service delivery models that focus on a student's

response to intervention (e.g., American Recovery and Reinvestment Act of 2009, ESSA, IDEA, etc.) allowing for Title I funds or other funding pools, such as School Improvement Grants (SIG) to support interventions. The school in which the present study is conducted receives both Title I and SIG funding to support the present tiered intervention program with allocations for materials, interventionists, and state-funded school support staff to assist in implementing interventions.

The issue of funding undoubtedly weighs heavily on the minds of policy makers and practitioners, yet other barriers to implementing and monitoring RtI programs exist. Policies aimed at supporting RtI implementation also require the use of research-based practices to be implemented at all tiers of intervention (e.g., NCLB, ESSA, IDEA). Yet, research on interventions remains limited. According to Strein et al. (2003) an examination of several leading school psychology journals demonstrated that only 3.3% of articles pertained to primary and secondary interventions, while also reporting practitioners and researchers alike ranked intervention articles as highest in priority. Gresham, MacMillan, Beebe-Frankenberger, and Bocian (2000) found only 14% of articles published in leading journals on learning disabilities were intervention studies. In assessing the volume of research on reading and math interventions, Seethaler and Fuchs (2005) found less than 6% of studies in school psychology and special education journals evaluated reading and math interventions. Furthermore, Burns et al. (2012) point to the fact that many of these articles fail to meet the *meets evidence standards*, or even *meets standards with reservation* designations established by the What Works Clearinghouse (WWC).

To compound the issue of a lack of sufficient methodological rigor in studies to guide specific intervention procedures, education practitioners report a lack of time or insufficient access to peer-reviewed studies and note the overwhelming task of evaluating such research (Bramlett et al., 2010). The issue of time to implement interventions is another challenge. This can be especially challenging in schools serving high concentrations of students in poverty and minority status where relatively large proportions of students are more likely to be at risk (Orfield, Ee, & Coughlan, 2017). While Gersten et al. (2009) suggest interventions should span 6 to 12 weeks, Codding et al. (2016) found that a dosage of 196 minutes of tiered instruction across four weeks produced significant and positive results. These barriers, and others not elucidated here, comprise the research-to-practice gap identified by many scholars and practitioners. One result of this gap is the need for practitioners to be able to evaluate a broad array of intervention protocols and procedures and to monitor systems of interventions across schools and school districts. The present study seeks to provide guidance for researchers and practitioners operating within the confines of educational contexts with limited fiscal resources, time, and access to peer-reviewed literature. Balancing the availability of resources, namely time (Castro-Villareal, Rodriguez, & Moore, 2014), with the need to maximize the impact of interventions (VanDerHeyden & Harvey, 2012) is of paramount concern in schools serving high concentrations of students at risk.

Framework for Systems Implementation. The AdvancED Standards *for* Quality Schools (2011) offer a lens for the application and integration of similar tiered intervention models in schools. AdvancED is a non-profit educational accreditation organization partnered with nearly 34,000 schools and school systems. AdvancED assesses schools and school districts with Standards *for* Quality for the purposes of accreditation and as part of diagnostic reviews – a school review system that uncovers root causes for underperformance and provides feedback to guide school improvement processes. Because AdvancED's Standards *for* Quality are known throughout a diverse range of schooling contexts, they provide a framework for understanding the implications of rigorously assessed tiered interventions and their relationship to equitable student learning opportunities as well as effective institutional improvement planning.

Conceptualizing Procedural Fidelity. We utilized methods to document procedural fidelity described by Barnett et al. (2014). The specific methods we utilized were: *self-report* (from the interventionist to one of the present researchers), *permanent product* (by gathering samples of student work) and *informal observations* (by visiting the intervention groups to observe that planned intervention activities were indeed being implemented with fidelity). Other methods described by Barnett et al. (2014) used to sample and monitor procedural fidelity include: direct observation, direct observation with a checklist, and the use of multiple methods. While direct observation is arguably the gold standard (Barnett et al., 2014), we selected multiple procedural fidelity sampling methods to monitor interventions as the present context show indications of high risk, as recommended by Barnett, Hawkins, and Lentz (2011) and Gresham (2009).

Conceptualizing Methodological Rigor. Ross et al. (2004) recommend two alternative methodological approaches with sufficient rigor when true experimental designs are not possible: multiple linear regression modeling and quasi-experimental designs. As recommended by Burns et al. (2012), we ensure methodological rigor for evaluating the efficacy of interventions through addressing the standards established by the What Works Clearinghouse (WWC, 2008) for quasi-experimental designs (QEDs). These standards include: (a) equating intervention and control groups, or matching participants on pre-tests; (b) overall and (c) differential attrition in the sample, as extreme attrition may bias any inferences drawn; (d) teacher-intervention confound; and (e) intervention contamination.

Critical Elements for Mathematics Interventions

Johnson, Mellard, Fuchs, and McKnight (2006) identify two fundamentally distinct forms of intervention in an RtI framework. The first is the standard treatment protocol approach, or a set of predetermined procedures that are evidence-based and have been found to be effective. A second approach is referred to as the problem-solving method, previously elaborated by Tilly (2002). Tilly (2002) describes the problem-solving method as an approach that focuses on finding supports to meet the identified needs of students through assessment, implementing evidence-based practices with fidelity, and continually monitoring the progress of students throughout the intervention as a means to inform potential instructional changes. Lembke et al. (2012) provide suggestions for instructional approaches to be utilized in mathematics interventions; notably the National Center on RtI (rti4success.org), practice guides from the What Works Clearinghouse (wwc.ed.gov), among others. Instructional and assessment strategies selected for use in our study were taken from a variety of sources, including resources cited by Lembke et al. (2012) as well as from mathematics educators working outside of the school that frames the context of our study. The design of each intervention was inductive and based upon identified student academic needs.

Methods

Our study is bound to a single high-minority/high-poverty urban public middle school in the Southeastern U.S., serving approximately 600 students in grades 6-8 during the 2015-2016 school year, all of which were male. Table 1 provides a summary of student demographics and other input variables as well as select school process variables. This school had never met adequate yearly progress (AYP) goals established by current state statutes, enacted in 2009, until the 2015-2016 school year which were largely attributed to increases in state-mandated mathematics test scores. The school was identified by the state education agency as a Persistently Low Achieving (PLA) school, or a school In Need of Improvement (INI). Under this designation, the school is eligible for and receives SIG funds to provide supplemental resources.

Table 1
Student Demographic Information and Other School and District Process Variables, 2015-2016

	Intervention (5.NBT.3)	Control (5.NBT.3)	Intervention (7.RP.1)	Control (7.RP.1)	School	District Middle School
No. of Students	12	10	26	25	584	20,709
Student Demographics						
% Black	58.3	60.0	46.1	56.0	41.8	37.4
% White	25.0	20.0	26.9	28.0	34.2	47.2
% Hispanic	-	20.0	19.2	8.0	15.9	8.7
% Other	10.0	-	7.7	8.0	8.0	6.7
% Free or Reduced-Price Lunch Participants	100.0	100.0	96.0	92.0	86.0	65.4
% English Language Learners	-	10.0	11.5	4.0	13.7	5.3
School Process						
% Teacher Attendance					95.5	95.1
% New Teachers					14.3	8.3
% Agree: Managing Student Conduct[a]					57.6	77.8
% Agree: Community Support[a]					61.9	79.5

Note. All students in the sample were male, as the school in which the study was conducted serves only male students. Categorizations of Race/Ethnicity are reported as the percent of students whose parents/guardians identify the students as one of the following: Black, Hispanic, White, Asian, Pacific Islander, American Indian, or 2 or More Races. % Other = the percentage of students self-identifying as a race/ethnicity excluding Black, Hispanic, or White. School process variables are included to provide some context for the school in which the study occurred. % Teacher Attendance = the average percent of active teachers counted as present or on professional leave across the school year. % New Teachers = the percent of teachers who are new to the school or intern teachers. % Agree: Managing Student Conduct and Community Engagement and Support are taken from the most recent Teaching, Empowering, Leading and Learning (TELL) survey results and are reported as measures of working conditions in the school. These process variables were selected because of their strong correlations to student achievement in the state in which the study occurred.
[a]Values reported are from the most recent school year for which data were available at the time of the study (2014-2015).

Through working with math teachers at this school, the need for implementing and monitoring multiple tiers of interventions to target identified student academic needs became apparent. Funds were available in the school's budget to hire a part-time academic interventionist. A retired middle school math teacher was identified and trained to conduct short-cycle tiered (pull-out) interventions in a resource room with students two days per week in the spring of 2016. An intervention design team consisting of the math interventionist and other school instructional staff worked cooperatively with grade-level math teams to identify essential standards for intervention, design pre- and post-tests, establish inclusion criteria based upon student pre-test scores, design bundled intervention lessons and formative assessment strategies, and evaluate data to gauge the

impact of the interventions based on a comparison of pre- and post-test scores between the students who received the intervention to a control group.

The Sample

Sampling procedures varied slightly from sixth to seventh grade because the sixth grade team elected to provide tiered interventions based on a fifth grade standard; a standard not explicitly taught in the sixth grade curriculum. Therefore, a single pre-test (universal screener) was given to all sixth grade students. The seventh grade team administered two pre-tests: a true pre-test and a second "pre-test" following the implementation of tier 1 instruction but prior to the implementation of the short-cycle interventions. The use of two pre-tests in the seventh grade allowed for the intervention efforts to focus on students for whom tier 1 instruction did not have a positive impact or had less of a positive impact, based on student test scores from the first pre-test to the second. Since the school in which the study was conducted served only male students, all of the students included in the sample were male.

Grade 6 Sample and Assignment. Since the sixth grade team elected to focus on a fifth grade standard (5.NBT.3) a single pre-test was given approximately two weeks prior to the start of the intervention. Students were identified for participation in the intervention program based on the level of need and the finite resources allotted to provide the interventions, selecting students at greatest risk (the lowest scores on the pre-test). Based on these criteria, N = 22 students were identified as needing intervention supports for all three learning objectives derived from the standard: n = 12 students were randomly assigned to the intervention group using an online random number generator (psychicscience.org/random.aspx) from an alphabetized list of students; the remainder (n = 10) were assigned to the control group. Students were placed into small (n = 4 to 6) intervention implementation groups based on their class schedules. Random assignment to small intervention implementation groups was not possible within the confines of the school day due to the insistence of school administrators and teachers that students not be pulled from their core content courses (English, math, science, or social studies). However, pre-test scores of students in the intervention treatment group and control group were compared to ensure baseline equivalence between the two groups. Descriptive statistics of student test scores are provided in Table 2 to demonstrate the level of matching that was achieved through the confines of student course schedules.

Grade 7 Sample and Assignment. Inclusion criteria to participate in the 7.RP.1 intervention were as follows:

1. Students who did not show growth from the tier 1 pre-test to the second pre-test, administered following tier 1 instruction.
2. Students who showed some growth toward meeting proficiency expectations, but had not yet met those expectations on the second pre-test, administered following tier 1 instruction.

A total of N = 51 students were identified who met the inclusion criteria for the seventh grade interventions. Students were first stratified based on the distinction identified in the inclusion criteria. Twenty-three students met the first inclusion criterion and 28 students met the second. The same online random number generator (psychicscience.org/random.aspx) was used to randomly select 13 students from an alphabetized list of students meeting the first criterion and another 13 students were selected who met the second criterion for placement in the intervention group (n = 26). The remaining n = 25 students were assigned to the control group and were not pulled for interventions but did take the post-test following the intervention with their peers assigned to the intervention group. The 26 students assigned to the intervention treatment were

placed into smaller intervention implementation groups, comprised of n = 4 to 6 students. As in the sixth grade sample, random assignment of students into smaller intervention implementation groups was not possible due to the confines of students' schedules. Descriptive statistics are provided in Table 2 for a comparison of pre-test scores for the intervention treatment and control groups.

Instrumentation

To measure the effect of interventions on student learning with respect to the identified standards, we created instruments using items released from prior state assessments. Intervention treatments and measures of mathematical competency are described below.

Dependent Variable: Measures of Mathematical Competency

Tests constructed by the intervention design team were used to measure students' competency with the identified mathematics standard. All tests consisted of ten selected response items pulled from state-released assessment items from prior state mathematics tests (Kentucky Department of Education, 2013). Test scores were reported as raw scores (i.e., the number of correct items). Proficiency expectations were made explicit prior to delivery of all assessments. Students failing to meet proficiency expectations on the pre-test were identified as possible participants for the tiered interventions. The pre-tests (which, in the case of the seventh grade group, was truly a post-tests of tier 1 instruction) served as universal screeners for identifying students needing interventions, as all students served in the general education math classroom of each grade level took the respective pre-test. The validity and reliability of the dependent variable is well-established since this study used state-released assessment items from prior state mathematics tests (Kentucky Department of Education, 2013).

Independent Variable: Intervention Treatments

Each intervention focused on a single standard and was conducted over the course of two weeks (implemented as four, one-hour intervention periods). Each grade-level math team identified an essential standard. The intervention team designed a series of four lessons using a bundle of instructional strategies in which models, manipulatives, and formative assessment practices were integrated in each intervention. Lessons were constructed based upon guidelines from Common Core (corestandards.org/Math) and WWC practice guides (wwc.ed.gov). These lesson plans were presented to the teachers on each grade-level math team such that critical feedback could be used to fine-tune the intervention plans prior to their implementation. The timing of these presentations was purposeful to avoid teacher-intervention confound and intervention contamination. What follows is a brief description of the standards and practices used in each intervention. While the intervention lessons were not entirely scripted, some parts of each lesson were. A sample lesson plan is provided in Figure 1.

7th grade – 7.RP.1	Day 3 of 4
Learning Target	→ I can use equivalent ratios to solve problems.
Establishing Engagement/Connection to Prior Knowledge	(7 – 10 minutes) Show a scale model of the Statue of Liberty. Teacher says: "This is a scale model of the Statue of Liberty… its scale is 1:129." Probing questions: "What do you think that means?" "What does the mean about the size of the REAL Statue of Liberty?" "How might we determine the height of the REAL Statue of Liberty, given this model?" (Materials: rulers, *Statue of Liberty* replica, calculators)
Deepening Understanding	(15 – 20 minutes) Discussion of Models: Similarity of shapes… when shapes are similar, you can infer the measurements of another if you know one. Provide scaffolding through gradual release of examples. (Guided notes: EquivalentRatiosofGeometricShapes.ppt, slides 1-5)
Applying Knowledge	(20 minutes) Scale model stations activity. (Materials: Scale models with prompts, 1 map with legend, rulers, calculators)
Assessment	(5 minutes) EquivalentRatiosofGeometricShapes.ppt (slide 10)

Figure 1. Sample daily lesson plan. Includes student-friendly learning target (objective). Systematic framing of the lesson can be seen in the progression by moving from the top toward the bottom beginning with Establishing Engagement through the culminating assessment of the lesson. Note the use of models and manipulatives throughout the lesson. Applying Knowledge provided the primary means through which error correction occurred through peer and teacher feedback.

Sixth Grade: Numbers and Operations in Base 10. The Common Core Mathematics Standard selected by the sixth grade math team for a targeted intervention was 5.NBT.3, "Read, write, and compare decimals to the thousandths" (Common Core, n.d., p. 35). While this standard is a fifth grade standard, the teachers of the sixth grade math team felt strongly about the need to structure tiered supports for this standard as they believed this standard inhibited many students from progressing toward meeting other standards in the sixth grade. A set of four, one-hour lessons were planned around three learning objectives: (a) students can translate numerical expressions between base-ten numerals, number names, and expanded form; (b) students can compare decimals to the thousandths based on meanings of the digits in each place; and (c) students can round numbers to the nearest place value. The intervention lasted four days spread out over two weeks (with students in each intervention group receiving two hours of instruction over two days each week). The post-test was administered the day immediately following the last day of the intervention. Each lesson was framed in a systematic fashion, used visual representations and other models of mathematical concepts, included purposeful and meaningful practice opportunities, and frequent progress monitoring and formative feedback (Bryant et al., 2011).

Seventh Grade: Ratios and Proportions. The Common Core Mathematics Standard selected by the seventh grade math team for a targeted intervention plan was 7.RP.1, "Compute unit rates associated with ratios of fractions, including ratios of lengths, areas and other quantities measured in like or different units" (Common Core, n.d., p. 48). The standard was collaboratively deconstructed by the intervention design team to ensure essential elements of this standard were represented in a learning progression. From this learning target progression, two learning objectives were selected to form the basis of the intervention: (a) the student can create equivalent

ratios and (b) the student can use equivalent ratios to solve problems. Two one-hour sessions were planned to provide intensive instruction around each of these learning objectives. Similar to the sixth grade intervention, the intervention lasted four days, spread out over two weeks with students in each intervention group receiving a total of four hours of instruction. The post-test was administered the day immediately following the last day of the intervention. Each lesson was framed in a systematic fashion, used visual representations and other models of mathematical concepts, included purposeful and meaningful practice opportunities, and frequent progress monitoring.

Research Design and Procedures

Each intervention was evaluated using a quasi-experimental design (QED) with control groups and pre-tests (Shadish, Cook, & Campbell, 2002). In one case, the case of the seventh grade students, two pre-tests were administered for purposes of identifying students. The two pre-tests were, in reality, a true pre-test (before) and a post-test for all students (after) tier 1 instruction in the general education classroom. These data help to isolate the impact of tier 1 and intervention treatments, reducing intervention confound (Koutsoftas, Harmon, & Gray, 2009) and allowed the intervention design team to use multiple baseline measures to identify students for possible inclusion in the intervention treatment or control groups through stratified random assignment. In our study, a post-test always refers to the test administered following the short-cycle targeted intervention, as the purpose of this study is to explicitly evaluate the efficacy of these interventions. In all cases, pre-tests were administered to all sixth and seventh grade students in attendance during the scheduled administration time. Students comprising the intervention treatment and control groups in attendance during the scheduled administration time were pulled out of class to take the post-test. Together, pre- and post-test scores for intervention treatment and control groups were used to test the hypothesis that the intervention treatment had a statistically significant and positive effect on student outcomes of a given mathematics standard.

Procedural Fidelity in Implementation. Several methods were used to document and ensure procedural fidelity of interventions, as suggested by Barnett et al. (2014) and Muñoz (2005). First, the interventionist participated in the intervention design process to ensure that she understood the instructional strategies and their purpose for their inclusion in each lesson. Second, the interventionist would self-report on the adherence to each day's lesson at the end of the day to the lead author. Noting Tilly's (2002) problem-solving intervention approach, some subtle changes were made in response to progress monitoring data and were documented by the interventionist on the lesson plan. These modifications were added to the lesson plan following a brief discussion on the justification of making each modification. Another method to monitor procedural fidelity was the collection of samples of student work completed in each of the small, intervention implementation groups. These products served a dual purpose, as they were also scored by the interventionist to monitor individual student progress and provide feedback to students about their learning. Finally, during each day that interventions occurred, the lead author would stop by for a brief informal observation (typically lasting about 5 minutes) to ensure the lesson activities planned were being implemented to fidelity. A plan of corrective action was in place to address any potential threat to procedural fidelity through verbal communication about the importance of adhering to the intervention plan as designed whenever necessary. No corrective action was needed throughout the duration of the interventions. However, some minor modifications were made to the lessons as, Tilly (2002) argues interventions must be responsive to student needs.

Statistical Analysis. GradeCam web-based software (gradecam.com) was used to gather student responses on all pre- and post-tests (which many teachers in the school commonly used to capture formative assessment data) and were later transferred to an electronic file using Statistical Package for the Social Sciences (SPSS) software (Version 22). Two members of the intervention design team independently checked the accuracy of this transfer of assessment data from GradeCam to the SPSS file. Data were always kept on a password protected computer and individual data files were encrypted with a password, providing an additional layer of data security. Individual student test scores were shared only with members of the intervention design teams. Following an analysis of descriptive statistics, a simple repeated measures design was employed for both interventions to compare differences between intervention treatment and control groups from the pre-test to the post-test. The *F* statistic was used to calculate the level of statistical significance. We assumed the nominal alpha criterion level, $\alpha = 0.05$. The partial eta-squared value was used to evaluate practical significance and the observed power was given attention in drawing causal inferences. The results for each intervention are reported below.

Results

This section reports the results of each intervention included in the study. Descriptive statistics of student outcomes on pre- and post-tests are reported in addition to inferential statistics computed from the repeated measures procedure in SPSS. Attention is given to both statistical and practical significance and the observed power is reported as well.

Grade 6: Number and Operations in Base 10

Descriptive Statistics. None of the students included in the intervention treatment group were absent from any of the intervention sessions. However, two students (both from the intervention treatment group) were unavailable to take the post-test following the intervention. These two students were removed from the data set. A total of N = 20 students were included in the analysis: n = 10 from the intervention treatment group and n = 10 from the control group. Descriptive statistics of student pre- and post-test scores are provided in Table 2.

Table 2
Descriptive Statistics of Student Pre- and Post-Test Scores 5.NBT.3 (N = 20)

	Mean Raw Score (Standard Deviation)	
	Pre-Test	Post-Test
Intervention Treatment (n = 10)	2.30 (1.06)	4.90 (2.23)
Control (n = 10)	2.50 (1.08)	3.30 (1.89)
Total (N = 20)	2.40 (1.05)	4.10 (2.17)

Note. The mean raw score is the average number of test items answered correctly on each test. Each test consisted of ten selected response items. Students not available for the post-test were excluded from the analysis.

Inferential Statistics. Box's test of equality of covariance matrices indicated that the assumption had been met ($p = 0.96$). A simple repeated measures analysis (Wilks' lambda) demonstrated a statistically significant difference between intervention and control groups from the pre-test to the post-test ($F[1, 18] = 18.58, p < 0.001$). The observed power was 0.58, indicating

that the sample size was sufficient to draw valid inferences (Stevens, 1996). Practical significance is also supported (partial eta-squared = 0.22). Figure 2 visualizes the estimated marginal means of student raw test scores on the pre- and post-tests between intervention treatment and control groups.
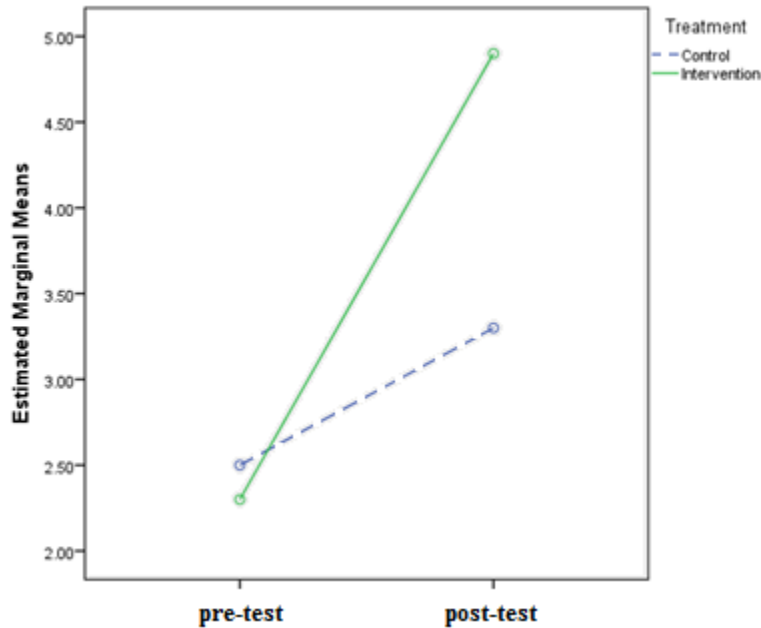

Figure 2. Estimated marginal means of student raw test scores for 5.NBT.3 intervention treatment and control groups on the pre-test and post-test.

Grade 7: Ratios and Proportions

Descriptive Statistics. Due to student absences and a limited time in which to capture post-test data, only N = 38 students (of the original 51) are included in the analysis of the seventh grade intervention, n = 17 students from the intervention treatment group and n = 21 students from the control group. Table 3 displays some descriptive statistics of pre- and post-test data. While both groups show growth from pre-test 1 to the post-test, the average growth of the intervention treatment group is larger than the average growth of the control group.

Inferential Statistics. An outlier ($z > 2.0$) was identified in the control group through preliminary exploration of the data. At the advice of Stevens (1996), the outlier was removed from the analysis leaving a sample size of n = 37. Box's test of equality of covariance matrices indicated that the assumption had been met ($p = 0.20$). A simple repeated measures analysis (Wilks' lambda) demonstrated a statistically significant difference between intervention and control groups from the pre-test to the post-test ($F[1, 35] = 5.67$, $p = 0.02$). The observed power was 0.64, indicating that the sample size was sufficient to draw valid inferences (Stevens, 1996). Practical significance was also supported (partial eta-squared = 0.14). Figure 3 visualizes the estimated marginal means of student raw test scores on the pre- and post-tests between intervention treatment and control groups.

Table 3
Descriptive Statistics of Student Pre- and Post-Test Scores 7.RP.1 (N = 38)

	Mean Raw Score (Standard Deviation)	
	Pre-test	Post-test
Intervention Treatment (n = 17)	3.18 (1.29)	5.47 (2.45)
Control (n = 21)	2.86 (1.79)	3.62 (2.18)
Total (n = 38)	3.00 (1.58)	4.45 (2.45)

Note. The mean raw score is the average number of test items answered correctly on each test. Each test consisted of ten selected response items. Students not available for the post-test or who were absent from intervention sessions were excluded from the analysis.
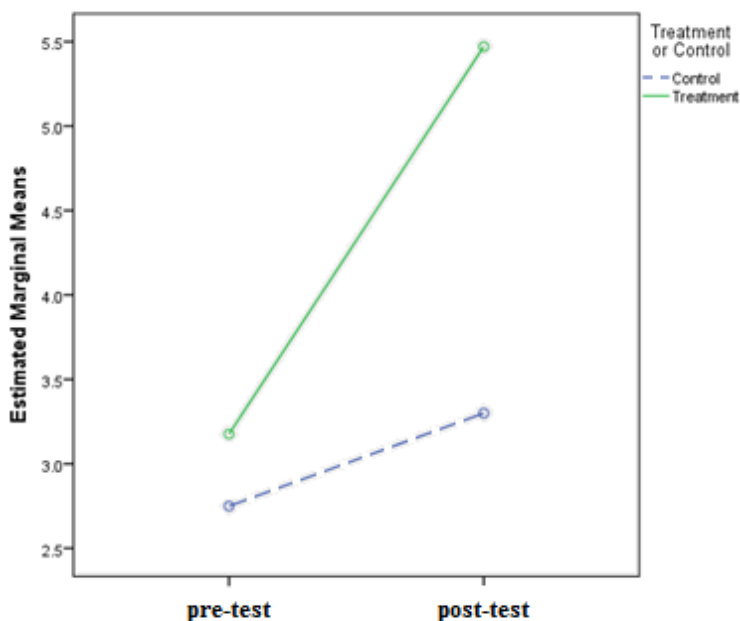

Figure 3. Estimated marginal means of student raw test scores for 7.RP.1 intervention treatment and control groups from pre-test to post-test.

Discussion

Our analysis of the efficacy of the short-cycle tiered mathematics interventions implemented in the present study indicated a significant (both statistically and practically) and positive impact on student learning as measured by pre- and post-tests, indicating the instructional and formative assessment techniques utilized herein may provide guidance to practitioners seeking to implement short-cycle, targeted interventions for the identified Common Core Mathematics Standards (5.NBT.3, 7.RP.1). These results support the inclusion of mathematical models and manipulatives with error correction (Lembke et al., 2012) and frequent formative assessment strategies to support student learning during small group interventions (Bryant et al., 2011; Wiliam, 2011). While some researchers have investigated the effect of dosage, frequency, or total

treatment duration on learning outcomes from tiered interventions (Codding et al., 2016; Denton et al., 2011; Yoder & Woynaroski, 2015), there is a paucity of research evaluating the efficacy of short-cycle mathematics interventions. The results of the present study support the findings of Codding et al. (2016) that short-cycle interventions can have a meaningful result on student learning. We hope that future research will explore the potential for short-cycle tiered interventions to meaningfully improve specific mathematical skillsets for students placed at risk.

It is important to consider the context of our study; practitioners and researchers are strongly encouraged to monitor the effectiveness of such interventions in other contexts. Indeed, we join Burns et al. (2012) in advocating for rigorous methodological approaches allowing for inferences of causality to be drawn in evaluating the efficacy of all interventions, behavioral and academic. We hope readers will use the results of rigorous methodological evaluation to inform the instructional approach of educators in all levels of tiered instruction and ensure all students receive the appropriate supports they need. We acknowledge the constraints of resources in many high-minority/high-poverty schools that can severely limit the amount of individualized tiered supports available to students (Orfield et al., 2017). Moreover, we acknowledge the limitations of educators to access peer-reviewed literature evaluating specific, tiered instructions when identifying opportunities to maximize the impact of limited resources (Bramlett et al., 2010). Nevertheless, when practitioners are forced to implement interventions rooted in tacit knowledge and personal experiences alone (Bramlett et al., 2002; Burns & Tsseldyke, 2009), we hope the present study provides a model to ensure rigorous evaluation.

Study Limitations and Implications for Future Research

Several limitations surface when considering the validity and reliability of the study. The ways students were identified for tiered supports varied across grade levels, as barriers to inclusion arose from forces both within and external to the school. These differences highlight the reality that practitioners must be dynamic in the design and implementation of numerous problem-based academic interventions. As we describe some of these nuanced limitations in the execution of the interventions, it is important to note two key components of the research design which cannot be removed from the design if arguments of causality are to be inferred: (a) the inclusion of at least one pre-test and (b) baseline equivalence of treatment and control groups (Shadish, Cook, and Campbell, 2002). These components are echoed as evidence standards for rigorous intervention evaluation established by the What Works Clearinghouse (2008).

One potential limitation was the lack of ability to randomly assign students in the intervention treatment group to small, intervention implementation groups. While we acknowledge the importance of random assignment of students into intervention implementation groups (Duhon, House, Hastings, Poncy, & Solomon, 2015; Yoder & Woynaroski, 2015), we also acknowledge temporal limitations on opportunities to provide tiered instructional services to students. The structure of the school's master schedule inhibited the ability to control for several potential covariates, such as time-of-day the treatment was administered, demographic variables such as race/ethnicity, and the possibility of exploring peer-effects between intervention implementation groups. Nonetheless, the research design does allow for a simple repeated measures analysis, thereby allowing for causal inferences to be drawn about the independent variable but does not allow for the inclusion of the identified covariates.

Lesson plans of seventh grade math teachers were reviewed to ensure intervention instruction was distinct from tier 1 instruction on the standards identified in the intervention program in an effort to minimize intervention confound (Koutsoftas, Harmon, & Gray, 2009). In

nearly all cases, direct and explicit instruction was given in tier 1 contexts and typically included a brief lecture followed by guided practice and then independent practice. While there are instances of direct and explicit instruction in the interventions, there was a focus on the use of mathematical visualizations of concepts and models; a characteristic absent from the lesson plans documenting tier 1 instruction of the associated standards.

As all students in the sample attend the same school and many have classes with one another in opposite groups (intervention and control), the diffusion of treatment remains a possibility (Creswell, 2014). Controlling for this threat to internal validity was not entirely possible in the context of our study, however the identities of members of the control group were never shared with the intervention group and the members of the control group were not notified of their participation until the day the post-test was administered. Compensatory threats (resentful demoralization and rivalry) were addressed by not informing students assigned to the control group about the intervention until after the experiment ended. A delayed treatment was administered to all members of each control group following the conclusion of the experiment, following the establishment of intervention efficacy through an analysis of the pre- and post-test scores. In addition, all students from both groups were given a reward for their participation in the post-tests (as recommended by Creswell, 2014).

Testing bias and instrumentation bias may have influenced student performance on the tests, which is why test items were modified and the answers for each item changed accordingly. In addition, several weeks typically passed between administering each test in a series. These measures were included in an attempt to strike a balance between testing and instrumentation bias (Creswell, 2014). Finally, while some may argue that the sample size of each intervention evaluated is small, special attention is given to the observed power. An explicit purpose of this study is to provide a methodological framework for evaluating short-cycle intervention programs within a single school, where sample size may be construed as a potential limitation in any school in which researchers and practitioners evaluate tiered interventions. We argue that a simple repeated measures analysis is robust with the sample size in our study, based upon our attention to the observed power in the statistical analysis. Practitioners and researchers alike should take caution in generalizing these results beyond contexts similar to those in our study.

Implications for Practice in Schools

AdvancED's Standards *for* Quality are adopted across a range of educational institutions and agencies. Standards *for* Quality exist for the assessment of schools, school systems, corporations, digital learning institutions, education service agencies, special purpose schools, and early learning schools (AdvancED, 2011). The school-based nature of the intervention under discussion in the current study dictates an analysis of the Standards *for* Quality for schools. There are five Standards *for* Quality for schools: Standard 1: Purpose and Direction; Standard 2: Governance and Leadership; Standard 3: Teaching and Assessing for Learning; Standard 4: Resources and Support Systems; Standard 5: Using Results for Continuous Improvement. Each standard contains a subset of indicators specifying each standard's conceptual aspects. Of particular importance to the current study are Standards *for* Quality three, Teaching and Assessing for Learning, and five, Using Results for Continuous Improvement. For the sake of concision in the current discussion, this analysis will not address individual indicators; rather, the focus of this discussion is directed at unifying concepts general to Standards *for* Quality three and five.

Standard *for* Quality three is paramount among the standards because of its focus on teaching, and more importantly, student learning. Key educational concepts present in Standard

for Quality three include: equitable student learning opportunities that ensure achievement of learning expectations, and instructional monitoring and the systematic adjustment of instructional practices (AdvancED, 2011). Ensuring that each student has equitable access to the achievement of rigorous learning expectations and that instructional systems and practices are adapted to better serve student learning needs requires a model for verifying the relationship between practice and learning. The research design and statistical procedures utilized herein to determine the efficacy of the math intervention provides a methodological framework (Shaddish et al., 2002; WWC, 2008) for adapting instructional practices to meet student learning needs embedded in evidence of learning (Tilly, 2002). Rigorously evaluating the impact of instruction on student learning throughout RtI's various tiers is necessary if educators are to fully monitor the learning environment and make informed systematic changes to the learning environment responsive to all student learning needs.

Standard *for* Quality five, Using Results for Continuous Improvement, includes concepts such as: comprehensive student assessment system, collecting and analyzing data from a range of sources, verifying improvement in student learning, and communicating comprehensive information about student learning and the conditions that support student learning (AdvancED, 2011). Similar to Standard *for* Quality three, the conceptual framework of standard five demands that quality schools and districts deploy a rigorous system to measure the impact of institutional decision making on student learning. Understanding the interrelationship between school improvement planning and student learning is essential to understanding the efficacy of educative practices, especially tiered intervention systems. Utilizing a rigorous methodological approach such as the one described in our study demonstrates implications for institutions to systematically inform the school improvement planning process with verifiable measures of the impact of decisions on student learning. Moreover, we acknowledge the importance of communicating comprehensive information not only about student learning impacted by tiered instructional supports, but about the conditions that support student learning. We join the calls from Barnett et al. (2014) and Barnett et al. (2011) to monitor the procedural fidelity of intervention implementation. We also echo the calls from Tilly (2002) to use measures of implementation fidelity to inform decisions about program adaptation to meet the needs of learners as well.

Conclusion

This study provides an empirical test of two, short-cycle tiered mathematic interventions implemented in a high-minority/high-poverty all-boys urban public middle school in the Southeastern U.S. and makes explicit the need for continued research on specific academic intervention strategies designed to meet the needs of all students, including short-cycle interventions. At the time of this study, we could not find any peer-reviewed journal articles explicitly tied to these Common Core Mathematics Standards. If the research-to-practice gap is to be mitigated, then researchers must heed the call for empirical validation of specific academic intervention practices and must ensure the adherence of the WWC standards for a rigorous methodological approach. In addition, practitioners must evaluate the fidelity of implementation and the efficacy of tiered academic supports to ensure limited resources are efficiently managed and justified.

References

AdvancED. (2011). Standards *for* Quality Schools. Alpharetta, GA: AdvancED. Retrieved from http://www.advanc-ed.org/sites/default/files/documents/SchoolStndsNolevels.pdf

Barnett, D.W., Hawkins, R.O., & Lentz, F.E., Jr. (2011). Intervention adherence for research and practice: Necessity or triage outcome? *Journal of Educational and Psychological Consultation*, *21*, 175-190.

Barnett, D.W., Hawkins, R.O., McCoy, D., Wahl, E., Shier, A., Denune, H., & Kimener, L. (2014). Methods used to document procedural fidelity in school-based intervention research. *Journal of Behavioral Education*, *23*, 89-107.

Bergan, J.R. (1977). *Behavioral consultation*. Columbus, OH: Charles E. Merrill.

Bergan, J.R., & Kratochwill, T.R. (1990). *Behavioral consultation and therapy*. New York: Plenum.

Bliss, S.L., Skinner, C.H., Hautau, B., & Carroll, E. (2008). Articles published in four school psychology journals from 2000 to 2005: An analysis of experimental/intervention research. *Psychology in the School*, *45*(6), 483-498.

Bramlett, R.K., Cates, G.L., Savina, E., Lauinger, B. (2010). Assessing effectiveness and efficiency of academic interventions in school psychology journals: 1995-2005. *Psychology in the School*, *47*(2), 114-125.

Bramlett, R.K., Murphy, J.J., Johnson, J., Wallingsofrd, L., & Hall, J.D. (2002). Contemporary practices in school psychology: A national survey of roles and referral problems. *Psychology in the School*, *39*(3), 327-335.

Bryant, D., Bryant, B.R., Roberts, G., Vaughn, S., Pfannenstiel, K., & Porterfield, J., et al. (2011). Early numeracy intervention program for first-grade students with mathematics difficulties. *Exceptional Children*, *78*(1), 7-23.

Burns, M.K., Klingbeil, D.A., Ysseldyke, J.E., & Petersen-Brown, S. (2012). Trends in methodological rigor in intervention research published in school psychology journals. *Psychology in the School*, *49*(9), 843-851.

Burns, M.K., & Ysseldyke, J.E. (2009). Reported prevalence of evidence-based instructional practices in special education. *Journal of Special Education*, *43*, 3-11.

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

Codding, R. S., VanDerHeyden, A. M., Martin, R. J., Desai, S., Allard, N., & Perrault, L. (2016). Manipulating treatment dose: Evaluating the frequency of a small group intervention targeting whole number operations. *Learning Disabilities Research and Practice*, *31*(4), 208-220.

Common Core State Standards Initiative. (n.d.). *Common core state standards for mathematics*. Retrieved from http://www.corestandards.org/wp-content/uploads/Math_Standards1.pdf

Davis, M., Herzog, L., & Legters, N. (2013). Organizing schools to address early warning indicators (EWIs): Common practices and challenges. *Journal of Education for Students Placed at Risk*, *18*, 84-100. doi: 10.1080/10824669.2013.745210

Dennis, M.S. (2015). Effects of tier 2 and tier 3 mathematics interventions for second graders with mathematics difficulties. *Learning Disabilities Research and Practice*, *30*(1), 29-42.

Deno, S. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, *52*, 219-232.

Denton, C. A., Cirino, P. T., Barth, A. E., Romain, M., Vaughn, S., Wexler, J., Francis, D. J., & Fletcher, J. M. (2011). An experimental study of scheduling and duration of "tier 2" first-grade reading intervention. *Journal of Research on Educational Effectiveness*, *4*, 208-230.

Duhon, G., House, S., Hastings, K., Poncy, B. C., & Solomon, B. G. (2015). Adding immediate feedback to explicit timing: An option for enhancing treatment intensity to improve mathematics fluency. *Journal of Behavioral Education*, *24*(1), 74-87.

Fletcher, J.M., Vaughn, S. (2009). Response to Intervention: Preventing and remediating academic difficulties. *Child Development Perspectives*, *3*(1), 30-37.

Fuchs, D., & Fuchs, L.S. (2006). Introduction to Response to Intervention: What, why, and how valid is it? *Reading Research Quarterly*, *41*(1), 93-99.

Fuchs, L.S., Fuchs, D., Craddock, C., Hollenbeck, K.N., & Hamlett, C.L. (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*(3), 491-509.

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J.R., & Witzel, B. (2009). Assisting students struggling with mathematics: Response to Intervention (RtI) for elementary and middle schools (NCEE 2009-4060). Washington, DC: 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/practiceguides

Gresham, F.M., MacMillan, D.L., Beebe-Frankenberger, M.E., & K.M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research and Practice*, *15*, 198-205.

Gresham, F.M. (2009). Evolution of the treatment integrity concept: Current status and future directions. *School Psychology Review*, *38*, 533-540.

Hill, D.R., King, S.A., Lemons, C.J., & Partanen, J.N. (2012). Fidelity of implementation and instructional alignment in Response to Intervention research. *Learning Disabilities Research and Practice*, *27*(3), 116-124.

Hoover, J.J., & Love, E. (2011). Supporting school-based response to intervention: A practitioner's model. *Teaching exceptional children*, *43*(3), 40-48.

Johnson, E., Mellard, D.F., Fuchs, D., & McKnight, M.A. (2006). *Responsiveness to intervention (RTI): How to do it*. Lawrence, KS: National Research Center on Learning Disabilities.

Kentucky Department of Education (2013). 2012-13 Technical Manual. Frankfort, KY: Author. Retrieved from https://education.ky.gov/AA/KTS/Documents/2012-13%20K-PREP%20Technical%20Manual%20v1%203.pdf

Klinger, J.K., & Edwards, P.A. (2006). Cultural considerations with response to intervention models. *Reading Research Quarterly*, *41*(1), 108-117.

Koutsoftas, A.D., Harmon, M.T., & Gray, S. (2009). The effect of tier 2 intervention for phonemic awareness in a response-to-intervention model in low-income preschool classrooms. *Language, Speech, and Hearing Services in Schools*, *40*(2), 116-130.

Lembke, E.S., Hampton, D., & Beyers, S.J. (2012). Response to Intervention in mathematics: Critical elements. *Psychology in the School*, *49*(3), 257-272.

Lembke, E.S., McMaster, K.L., & Stecker, P.M. (2010). The prevention science of reading research within a response-to-intervention model. *Psychology in the School*, *47*(1), 22-35.

Mellard, D., & McKnight, M., & Jordan, J. (2010). RTI tier structures and instructional intensity. *Learning Disabilities Research and Practice*, *25*(4), 217-225.

Munoz, M. A. (2005). Black box. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 34-35). Thousand Oaks, CA: Sage.

Orfield, G., Ee, J., & Coughlan, R. (2017). *New Jersey's segregated schools: Trends and paths forward.* Los Angeles: University of California, UCLA Civil Rights Project.

Riccomini, P.J., & Witzel, B.S. (2010). *Response to intervention in math*. Thousand Oaks, CA: Corwin Press.

Ross, S. M., Nunnery, J. A., Goldfeder, E., McDonald, A., Rachor, R., & Hornbeck, M. (2004). Using school reform models to improve reading achievement: A longitudinal study of Direct Instruction and Success For All in an urban district. *Journal of Education for Students Placed At Risk*, *9*, 357–388.

Sanetti, L.M.H., Gritter, K.L., & Dobey, L.M. (2011). Treatment integrity of interventions with children in the school psychology literature from 1995 to 2008. *School Psychology Review*, *40*, 72-84.

Sanetti, L.M.H., & Kratochwill, T.R. (2009). Toward developing a science of treatment integrity: Introduction to the special issue. *School Psychology Review*, *38*, 445-459.

Seethaler, P.M., & Fuchs, L.S. (2005). A drop in the bucket: Randomized controlled trials testing reading and math interventions. *Learning Disabilities Research and Practice*, *20*, 98-102.

Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs: For generating causal inference*. Belmont, CA: Wadsworth, Cengage Learning.

Sirin, S. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*(3), 417-453.

Skinner, C.H., & Skinner, A.L. (2007). Establishing an evidence base for a classroom management procedure with a series of studies: Evaluating the color wheel. *Journal of Evidence-Based Practices for Schools*, *8*, 88-101.

Strein, W., Cramer, K., & Lawser, M. (2003). School psychology research and scholarship. *School Psychology International*, *24*, 421-436.

Tilly, W.D. (2002). Best practices in school psychology as a problem solving enterprise. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology* (4th ed., pp. 21-36). Bethesda, MD: National Association of School Psychologists.

White, K. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, *91*, 461-481.

VanDerHeyden, A. M., & Harvey, M. (2012). Using data to advance learning outcomes in schools. *Journal of Positive Behavior Interventions*, *15*, 205–213.

Vaughn, S. R., Wanzek, J., Woodruff, A.L., & Linan-Thompson, S. (2006). A three-tier model for preventing reading difficulties and early identification of students with reading disabilities. In D. H. Haager, S. Vaughn, & J. K. Klingner (Eds.), *Validated reading practices for three tiers of intervention* (pp. 11-28). Baltimore, MD: Brookes.

What Works Clearinghouse. (2008). *Evidence standards for reviewing studies (Version 1.0)*. Washington, DC: Institute of Educational Sciences.

Wiliam, D. (2011). *Embedded formative assessment*. Bloomington, IN: Solution Tree Press.

Yeaton, W.H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology*, *49*, 156-168.

Yoder, P. J., & Woynaroski, T. (2015). How to study the influence of intensity of treatment on generalized skill and knowledge acquisition in students with disabilities. *Journal of Behavior Education*, *24*(1), 152-166.

Zirkel, P.A., & Thomas, L.B. (2010). State laws and guidelines for implementing RTI. *Teaching Exceptional Children*, *43*(1), 60-73.