Brian R. Bryant
Diane Pedrotty Bryant
Greg Roberts
Anna-Mária Fall
The Meadows Center for Preventing Educational Risk, The University of Texas at Austin

Effects of an Early Numeracy Intervention on Struggling Kindergarteners' Mathematics Performance

Abstract

The purpose of this study was to investigate the effects of an early numeracy intervention delivered by kindergarten teachers to students identified as having mathematics difficulties. A multigroup growth-modeling-with-random-assignment-to-intervention-condition design was employed. Thirty-two teachers were randomly assigned to the treatment or comparison condition. A total of 71 students participated in the study, 47 in the treatment group and 24 in the comparison group. Results indicated that the treatment condition students outperformed comparison students ($g^* = .99$) and demonstrated statistically significantly higher scores on all proximal measures of early numeracy. Also, about 80% to 100% of the variance was accounted for at the student level. Performance on distal measures was less impressive, with no significant differences between groups; the effect size was .44. Teachers rated components of the intervention highly, reflecting strong teacher satisfaction.

Researchers have recently demonstrated that early mathematics achievement at the kindergarten level to some extent predicts later academic performance (Jordan, Kaplan, Ramineni, & Locuniak, 2009). For example, Morgan, Farkas, and Wu (2009), in their analyses of the Early Childhood Longitudinal Study-Kindergarten Cohort (ECLS-K; National Center for Educational Statistics [NCES, 2004]) database, found that the learning trajectories of students with mathematics problems in the fall and spring of kindergarten continued to show slower mathematical growth throughout their early academic careers.

These findings do not mean that children who are slow to learn early mathematics skills are destined to struggle throughout their academic careers. However, given an overall persistent pattern of low mathematics performance for some students compared to typically achieving students (NCES, 2015), it is not surprising that students who struggle with early mathematics number sense attainment have difficulty obtaining more complex mathematical skills and concepts (Stinson, 2004; Wu, 2001).

Although some students who enter kindergarten demonstrate an adequate understanding of early number concepts, others lack the informal knowledge of mathematics that contributes to primary level mathematic success (Bryant et al., 2011). According to Wilson, Revkin, Cohen, Cohen, and Dehaene (2006), for example, children with mathematics difficulties exhibit a variety of fundamental mathematical problems, including difficulties with counting, recalling arithmetic facts, and representing quantity and/or linking quantity to symbolic number representations. Because early mathematics ability is predictive of later achievement, early intervention in numeracy concepts and skills should be available for all students who exhibit mathematics difficulties (Bryant et al., 2011; Morgan et al., 2009).

Bryant, B. R., Bryant, D. P., Roberts, G., & Fall, A. (2016). Effects of an early numeracy intervention on struggling kindergarteners' mathematics performance. *International Journal for Research in Learning Disabilities, 3*(1), 29-45.

Research on the Development of Number Sense Interventions

Although lagging behind early reading intervention research, in recent years, early numeracy interventions, which incorporate practices to promote number sense (i.e., fostering an understanding of the "conceptual relationships between quantities and numerical symbols" [Griffin, 2004, p. 39]), have become more prevalent for younger students. For example, in one study, Clarke et al. (2011) compared the performance of 56 kindergarten students who participated in a combination of core instruction and a supplemental intervention (ROOTS) to peers (n = 64) who participated in core instruction only. Results on the *Test of Early Mathematics Ability-3* (TEMA-3, Ginsburg & Baroody, 2007) demonstrated that students receiving the combined approach (i.e., core plus supplemental instruction) made significantly greater gains than those receiving core instruction only. Further, students identified as being at risk (i.e., those performing below the 40th percentile on the TEMA at pretest) in the treatment group significantly outperformed their at-risk comparison counterparts on the TEMA ($t = 3.29$, $p = .0017$, $g = .24$). In a more recent study (Clarke et al., 2016) with a larger sample (29 kindergarten classrooms randomly assigned to the treatment [ROOTS] or comparison condition), similar findings were noted, with Hedges' g effect sizes of .38 for the TEMA standard score and .30 for the *Early Numeracy-Curriculum Based Measurement* (Clarke & Shinn, 2004).

In another study, Jordan, Glutting, Dyson, Hassinger-Das, and Irwin (2012) examined the effectiveness of a targeted small-group number sense intervention for high-risk kindergarteners from low-income families. Children were randomly assigned to one of three participant groups. The first group received a number sense intervention; the second group received a language intervention; and the third group was a business-as-usual comparison group. Interventions lasted 30 minutes each and were administered three times per week; in all, 24 lessons were taught. The intervention and comparison groups each were taught the same core mathematics curriculum; during the mathematics intervention, comparison students engaged in typical kindergarten activities such as learning centers or special subjects instruction.

Immediate posttest results on the Number Sense Brief (NSB) (Jordan, Glutting, Ramineni, & Watkins, 2010) revealed that the group that received the number sense intervention performed better than the comparison group, with positive effect sizes (most Cohen ds were medium to large) on measures of number sense and general mathematics. Further, following the administration of the NSB and the *Woodcock-Johnson III Tests of Achievement* (WJ) *Form C Brief Battery: Applied Problems and Calculation* subtests (Woodcock, McGrew, Schrank, & Mather, 2007) eight weeks after the intervention concluded, results remained similarly positive. According to the authors, no significant differences were found between language and comparison groups on either of the mathematics measures.

In a more recent study, Dyson, Jordan, Beliakoff, and Hasinger-Das (2015) randomly assigned 126 kindergarten children to one of three groups. The first group (n = 44) was administered a number sense intervention followed by a number fact practice session. The second group (n = 40) received the same number sense intervention, followed by a number list practice session. Finally, the third was a business-as-usual comparison group (n = 42). The interventions were delivered in a small-group setting over twenty-four 25-minute lessons for an eight-week period. The comparison group received 30 minutes of mathematics time to supplement their core instruction. Immediately following the intervention, and eight weeks later, measures of number sense, arithmetic fluency, and general mathematics calculation achievement were administered. Results showed that although not always significantly different, the treatment groups performed better than the comparison groups based on a computed Hedges g of greater than .25, which is viewed as an indication of effective educational practice (What Works Clearinghouse, 2014; http://ies.ed.gov/ncee/wwc/pdf/wwc_version1_standards.pdf).

Thus, evidence is mounting regarding the ability to improve the number sense performance of kindergarten students who are at risk for mathematics difficulties. Given the growing evidence of effective early numeracy interventions, the essential components of such practices should be examined.

Essential Features of Early Numeracy Interventions

Essential features of intervention have been identified and incorporated into research protocols as a result of an increasing number of studies on early numeracy improvement for struggling students and informed by previous key syntheses of effective mathematics practices (e.g., Gersten, Beckmann et al., 2009; Gersten, Chard et al., 2009).

Specifically, researchers have recommended that prevention and intervention efforts include verbalizations of cognitive strategies (Fuchs & Fuchs, 2001), as well as physical (concrete) and visual (pictorial) representations of number concepts (Bryant et al., 2011, 2014; Fuchs & Fuchs, 2001; Gersten, Chard et al., 2009).

In addition, findings from studies on students with mathematics difficulties support the use of explicit, strategic instruction in procedural knowledge and conceptual understanding, such as the commutative property of addition and counting strategies (Baker, Gerstein, & Lee, 2002; Bryant, 2005; Clarke et al., 2016; Gersten, Jordan, & Flojo, 2005). Notably, practices such as cognitive modeling/thinking aloud make transparent for struggling students ways of approaching difficult problems through the use of visual representations and precise mathematical language. All students, and especially those with mathematics difficulties, can benefit from ample opportunities to practice (e.g., guided practice, distributed practice) using visual representations (e.g., number lines, manipulatives, mathematics models) and mathematics vocabulary, as modeled by their teachers, to discuss concepts and solve problems in small groups and individually (Greenes, Ginsburg, & Balfanz, 2004). Additionally, effective teachers use questioning strategies to help students connect representations and verbal and symbolic statements (Clements & Sarama, 2004).

The Present Study

As noted, studies have provided compelling evidence showing that young children's early numeracy understanding can be improved and supporting the importance of teachers at the primary level focusing on mathematics foundation skills that set the stage for later mathematics success. Yet, more research is needed to inform the field about effective practices to enhance the performance of young students who manifest difficulties with mathematical concepts and number sense. Not only should efforts be made to increase the performance of concepts and skills targeted for interventions, evidence is also needed to show whether students can generalize or transfer this knowledge to broader mathematics assessments that measure overall mathematics achievement.

Finally, it is important to understand teachers' perspectives, or satisfaction, about the interventions they are teaching. According to Wolf (1978), *social validity* refers to the extent to which participants (i.e., the teachers in this study) delivering behavioral and academic interventions find them acceptable in terms of the goals of the intervention, the appropriateness of the procedures, and the importance of the treatment implications.

One would be hard pressed to dispute the claim that, for young students, demonstrating improvement in critical academic areas is "socially important." As Cooper, Heron, and Heward (2007) observed, social validity involves efforts (e.g., interventions) that can make a socially significant change in an individual's life. Thus, the degree to which teachers attribute improved student performance to a specific intervention indicates their perspective about the social validity of the intervention lessons.

The purposes of the present study were threefold: (a) To present the findings on the effects of a supplemental intervention on performance on an early numeracy measure; (b) To determine the effects of the intervention on subjects' performance on a mathematics achievement test that broadly measured mathematics concepts and skills; and (c) To determine the degree to which teachers perceived the early numeracy intervention as being socially valid.

The following three research questions guided the study:

1. Did students receiving the early numeracy Tier 2 supplemental intervention demonstrate improved performance on progress monitoring measures of early numeracy mathematics compared to students receiving business-as-usual mathematics instruction with no particular intervention? We hypothesized that students in the treatment group would outperform students in the business-as-usual comparison group.

2. Did students receiving the early numeracy Tier 2 intervention demonstrate improved performance on a distal standardized measure of a broad array of mathematics concepts and skills compared to students receiving business-as-usual mathematics instruction? We hypothesized that there would be no differences between groups on the distal measure because we did not directly teach the skills and concepts measured on the test.
3. To what degree was teacher satisfaction observed on the teacher survey of the practices contained in the lessons? We hypothesized that the treatment teachers' satisfaction would be high, on average, as a result of the specific instructional design and delivery practices that included in the lessons.

Method

Participants and Setting

A total of 71 kindergarten students participated in the study. Students attended 16 schools in Texas and were participating in a statewide response to intervention (RtI) supplemental early mathematics intervention. The schools were located in major metropolitan, urban districts.

At each school, classroom teachers administered an early mathematics measure, the *Texas Early Mathematics Inventories - Progress Monitoring* (TEMI-PM; Texas Education Agency/University of Texas System [TEA/UTS], 2008a), used for universal screening purposes, to students from intact classrooms who had received institutional review board (IRB)-approved consent to participate in the study were. Students qualified for the study if they received a total score below the 25th percentile, based on Texas statewide normative data. All students received mathematics instruction in the general education classroom; students who received mathematics instruction in a bilingual classroom were not part of the study. Because of their young age, none of the students had been identified as having a learning disability but were identified as having mathematics difficulties based on the results of the screener (TEMI-PM, 2008a). Table 1 shows the demographic characteristics of the participants who completed the intervention.

Table 1
Participant Demographics for Treatment and Comparison Groups

Demographic	Total *N* = 71	
Category	Treatment *N*	Comparison *N*
Gender		
Female	23	7
Male	24	16
Missing	-	1
Ethnicity		
African American	6	4
Hispanic	30	2
Caucasian	11	15
Missing	-	3
Free/Reduced-Price Lunch		
Yes	27	18
No	20	3
Missing	-	3
EL/LEP		
Yes	12	3
No	35	19
Missing	-	2

ELL = English Language Learners; LEP = Limited English Proficient; N/A = Information not available.

Teachers

A total of 32 teachers, 16 treatment and 16 comparison, across 16 schools participated in the study. All teachers were female, were certified by the state of Texas to teach kindergarten classes, and were deemed highly qualified according to *No Child Left Behind* reporting criteria, in effect at the time of the study. Six teachers held master's degrees; two held teaching certification through alternative licensure; the remaining teachers held bachelor's degrees and were certified through their college/university teacher certification program; one teacher was African American; the remaining teachers were White.

As a project designed to validate the intervention and assessments for use in programs such as RtI, classroom teachers administered all tests and implemented the interventions. Not all RtI programs use teachers as interventionists; however, in this study teachers were responsible for testing and the delivering the intervention to ensure that interventionists were certified teachers who knew the students well.

Research Design

A multigroup growth-modeling-with-random-assignment-to-intervention-condition design was employed. Teachers were randomly assigned to the treatment or comparison condition. Of the 71 students participating, 47 were in the treatment group and 24 in the comparison group. The approximately two-to-one ratio was established in negotiations with participating school principals, who requested that as many students as possible be assigned to the treatment group (Bryant et al., 2011). Three students (all from the intervention group) were lost to attrition during the course of the intervention.

Measures

TEMI-PM (TEA/UTS, 2008a). The TEMI-PM is a researcher-developed measure that was commissioned by the state education agency for use by Texas teachers. The TEMI-PM includes versions for kindergarten and first and second grade, all of them standardized on more than 1,700 students at each grade level across the state of Texas. The kindergarten version consists of three alternate, equivalent forms that are administered in the fall, winter, and spring, respectively. Each form is composed of four 2-minute timed subtests that are group-administered and involve numbers ranging from 0 through 20 to conform to Texas standards for instructional content.

The first subtest, *Magnitude Comparison*, contains 64 items. Students look at two numbers that appear side-by-side in a box in their student booklet (a vertical dotted line separates the two numbers) and are given 2 minutes to circle the bigger of the two numbers, or both numbers if they are the same. Alternate-forms reliability ranges from .76 to .85, with a median of .81. Similar measures are available online (www.interventioncentral.org/htmdocs/interventions/cbmwarehouse.php). The test is similar to Clarke and Shinn's (2004) *Quantity Discrimination Measure Verbal 1-20*.

For *Number Identification*, the second subtest, students look at rows and columns of squares and count the number of squares shown. Students then circle from four response choices the answer that shows "how many." The 2-minute subtest includes 28 items. Reliability coefficients range from .75 to .83, with a median of .78. Number identification tasks have appeared on many early mathematics tests (e.g., *KeyMath-3* Connolly, 2008; *Woodcock-Johnson Revised*, Woodcock, McGrew, Schrank, & Mather, 2007; and Clarke and Shinn's, 2004, *Number Identification Measure 1–20*).

Number Sequences, the third subtest, contains 42 items. Students look at a three-number sequence; one number of the "counting by ones" sequence is missing and is represented by a blank. The missing number may be either the first, second, or third number of the sequence. Students are given 2 minutes to look at four response choices and circle the one that represents the missing number. Reliabilities range from .81 to .85, with a median of .84. The skill of identifying missing numbers in a sequence is often found on tests that assess number sense. Also, missing number tests are available online (http://www.interventioncentral.org/curriculum-based-measurement-reading-math-assesment-tests); also, Clarke and Shinn (2004) include *Missing Number Measure Blank Varied 1-20* as one of their early mathematics screening measures.

For the fourth subtest, *Quantity Recognition*, students look at randomly placed dots, ranging from one to six dots, clustered near one another and circle the response choice (1, 2, 3, 4, 5, or 6) that corresponds to the number of dots shown. Students are given 2 minutes to complete as many of the 70 items as they can. Reliabilities range from .72 to .84, with a median of .80. Kaufman, Lord, Reese, and Volkmann (1949) used the term *subitizing* to describe the rapid, correct, and self-assured judgment of the quantity represented by small numbers of items. Several researchers, including Benoit, Lehalle, and Jouen (2004), have used formats similar to *Quantity Recognition*.

Finally, the Total Score is calculated by summing the raw scores of all four subtests. The alternate-forms reliability coefficients for the kindergarten TEMI-PM Total Score range from .88 to .92, with a median of .89. In addition, area under the Receiver Operating Characteristic (or ROC curve) values for the fall form of the TEMI-PM Total Score was.80, demonstrating good predictive power (Minitab, n.d.) for the criterion measure, the *Stanford Achievement Test* (10th ed.) (SAT-10; Harcourt Assessment, 2003).

Texas Early Mathematics Inventories-Outcome (TEMI-O; TEA/UTS, 2008b). The TEMI-O was co-normed with the TEMI-PM. The kindergarten version consists of a single subtest, *Mathematics Problem Solving* (MPS), which asks students to respond to 34 items that assess a broad array of mathematics concepts and skills (e.g., counting and cardinality, number and operations, geometry, measurement and data, National Council of Teachers of Mathematics [NCTM], 2006; National Governors Association Center for Best Practices and Council of Chief State School Officers [NGACBP/CCSSO], 2010). The teacher provides a stimulus prompt (e.g., "Look at the first box. Rasheeda was given two big buttons and one little button by her aunt. Her aunt asked her, 'How many buttons do you have now?' Now look at the other boxes. Mark the box that shows how Rasheeda used a ten frame to find the answer."), and students are to select from three response choices the one that provides the answer to the prompt. Coefficients alpha across forms reported in the TEMI-O Technical Manual (TEA/UTS, 2008c) for kindergarten for Forms A, B, and C, respectively, were .87, .84, and .83.

Prior to administering the measures to students in the fall, members of the research team went to schools to train participating kindergarten teachers on the administration of the TEMI-PM and the TEMI-O. Upon completion of each testing cycle (fall, winter, spring), teachers turned in the test protocols to the school liaison for the project, who mailed them to the project coordinator. There were approximately three months between each assessment time period: fall, winter, and spring. For each administration of the TEMI-PM and TEMI-O, a commercial data entry company entered item responses into a data file on computers (guaranteed 98% accuracy), and the lead author electronically scored items in preparation for data analyses.

Teacher satisfaction survey. A teacher satisfaction scale was constructed consisting of 18 statements that focused on content of the lessons, alignment of the content with state standards, procedures for teaching the lessons, and benefit for students. The items were chosen based on Wolf's (1978) definition of socially validity (i.e., goals of the intervention, the appropriateness of the procedures, and the importance of the treatment implications). The survey was designed in an electronic, online platform (https://surveystation.austin.utexas.edu); teachers responded to each of the items using a Likert-type scale with ratings ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). Internal consistency reliability for the teachers' social validity scale was .77.

Fidelity of Implementation

Fidelity, which refers to the act of measuring how well an intervention is being implemented compared to the original intervention design, is a vital component of an intervention study. The project coordinator and the school liaison conducted the fidelity observations. The coordinator conducted the first observation in November, which was one month following the training workshop. The school liaison conducted the second observation in December, and the coordinator conducted the third observation in February or March. Prior to conducting the second fidelity observation, the school liaison attended a webinar on fidelity observation and read a brief article on the importance of fidelity as part of an intervention.

In all, each teacher was observed for 3 sessions for the 22-week dosage of the intervention to assess adherence to the intervention. Quality performance indicators included (a) following the scripted lessons for the content (e.g., uses manipulatives and teaches strategies as stated); (b) implementing the instructional procedures (e.g., reviews background knowledge, provides modeling, provides corrective feedback); and (c) managing student behavior and materials (e.g., obtains and maintains student attention, intervenes quickly to redirect behavior). Performance indicators were rated on a 0- to 4-point scale, in which 0 = Very Poor (Adhered very little or not at all to scripted instructions), 1 = Poor (Adhered to some but not much to scripted instructions), 2 = Fair (Adhered to much of the scripted instructions) 3 = Good (Adhered mostly to scripted instructions), 4 = Excellent (Adhered perfectly to scripted instructions).

Mean overall results across all teachers showed the following: 1st observation: 3.00, 2nd observation: 4.00, and 3rd observation: 3.40. After each observation, results were shared with the teachers, and suggestions for improvement, if any were needed, were discussed. Overall results across grades and times showed a moderate to high degree of fidelity in the implementation of the lessons.

Procedures

Professional development and just-in-time resources. Researchers provided a full day of professional development (PD) for the interventionists and each school's liaison (e.g., counselor, lead grade-level teacher). Participating kindergarten teachers were brought to a central training location. The training included a review of the lessons and materials, an opportunity to view a video of teachers implementing the lessons, and time to practice the lessons under the guidance of the workshop leaders. The interventionists were also given the lessons and accompanying materials, including copies of the student booklets. In addition, training was provided on how to administer the TEMI-PM and TEMI-O, which the teachers were responsible for administering. The schools' liaisons, in turn, were responsible for providing TEMI-PM and TEMI-O administration instructions to the comparison group teachers using the training procedures utilized during the full day of PD.

When they returned to their schools to start the intervention, just-in-time resources were made available to treatment teachers and school liaisons. These resources included webcasts or newsletters. Webcasts were conducted on topics (e.g., progress monitoring) that the project coordinators and school liaisons thought were important based on their site visits and classroom observations, respectively. A project coordinator from the research team visited the interventionists twice during the school year. Coaching was provided by the project coordinator as further PD as needed to each teacher. These visits also included a fidelity check and meeting with the project's liaison or principal.

Early numeracy intervention. Teachers in the treatment group were responsible for teaching the supplemental intervention during their already designated small-group (three or four students) instruction time. Intervention dosage consisted of 4 days per week for 25 to 28 minutes over the course of 23 weeks.

The supplemental intervention was developed with an emphasis on early numeracy concepts and skills. Specifically, the curriculum included identifying and writing numerals; counting, ordering, and comparing quantities; identifying part-part-whole quantities; making groups; and solving simple change problems with the result unknown. Visual representations were used as scaffolds to help students develop and build knowledge of concepts, operations, and properties (e.g., commutative property, associative property). Specifically, the visual representations were intended to help students construct connections between mathematical concepts by using manipulatives (e.g., connecting cubes, base-10 materials), pictorial representations (e.g., 10 frames, dot configurations for facts, place-value models), and symbolic representations, which are critical components of conceptually based instruction (Baroody, 1990; Clements & Sarama, 2009; Gersten, Beckman, et al., 2009; Hiebert & Wearne, 1992; NCTM, 2006; Sarama & Clements, 2009). Vocabulary development was also an important aspect of the intervention. A glossary with words and their definitions was included to help teachers as they taught the mathematical content.

The instructional design and delivery of the lessons consisted of the critical features of explicit instruction that have been validated in numerous studies with struggling students (e.g., Bryant et al., 2011; Fuchs et al., 2001; Swanson, Hoskyn, & Lee, 1999). The features included a teaching routine consisting of modeling, guided practice, and independent practice (progress monitoring); error correction procedures; pacing; opportunities for meaningful practice (e.g., with visual representations); examples; and review.

Finally, daily activity level progress monitoring was conducted during independent practice at the end of each lesson. Teachers gave students four oral or written problems to determine their response to instruction on each lesson; that is, whether they met the lesson's objective. Students had to demonstrate accuracy on three out of four of the problems to consider the lesson successful for that student.

Comparison classroom practices. Sixteen teachers served as interventionists for the comparison groups. Classroom practices for the comparison group varied, depending upon the school. Yet, all programs focused on kindergarten Texas Essential Knowledge and Skills (TEKS; http://ritter.tea.state.tx.us/rules/tac/chapter111/ch111a.html) content, which involves skills and concepts in number and operations. Examples of instructional content for the TEKS include (a) counting forward and backward to at least 20 with and without objects; (b) reading, writing, and representing whole numbers from 0 to at least 20 with and without objects or pictures; (c) counting a set of objects up to at least 20 and demonstrating that the last number said tells the number of objects in the set regardless of their arrangement or order; and (d) recognizing instantly the quantity of a small group of objects in organized and random arrangements.

Most comparison group teachers provided additional supportive instruction for the lessons that were being taught daily to the whole class. A combination of explicit, systematic instruction and inquiry-based instruction was typically used, although the extent to which each was employed varied by teacher. In short, the content taught was fairly consistent across schools, but the manner with which information was presented varied. Very little progress monitoring (i.e., record keeping of consistent formative evaluations) was observed.

Results

Descriptive results for treatment and comparison classrooms at pretest and posttest are summarized in Table 2. As illustrated, no statistically significant pretest differences were found between students in the intervention and comparison groups, suggesting pretreatment equivalence on all of the measures, including Magnitude Comparison (β = -1.11, *SE* = .63, *p* = .08; β – standardized coefficient), Number Identification (β = -.05, *SE* = .56, *p* = .93), Number Sequence (β = -.12, *SE* =.36, *p* = .74), Quality Recognition (β = -.83, *SE* = 1.55, *p* = .57), TEMI-PM Total Score (β = -2.43, *SE* = 2.01, *p* = .23), and the TEMI-O Mathematics Problem Solving (β = 1.12, *SE* = 1.53, *p* = .26).

To evaluate the statistical significance of group differences, we fit multilevel models (MLwin 2.10; Rasbash, Steele, Browne, & Goldstein, 2004), nesting students in classes and estimating effect at the classroom level (see Table 3). Family-wise error associated with multiple comparisons was controlled using the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995), and effect sizes were estimated as the ratio between the model-derived treatment coefficients and the unadjusted pooled within-group standard deviation across conditions at posttest (i.e., Hedges *g* with small sample correction, henceforth indicated as Hedges *g**). The regression coefficients represent the performance difference between treatment conditions at posttest, controlling for pretest differences.

Effects on the Early Numeracy Progress Monitoring Measures

The kindergarten treatment group outperformed the comparison group on Magnitude Comparison (β = 11.39; *g** of .73), Number Identification (β = 3.77; *g** = .95), Number Sequences (β = 5.46; *g** = .76), Quantity Recognition (β = 9.77; *g** = .96), and the TEMI-PM Total Score (β = 31.97; *g** = .99). Approximately 84% of the total variance was at the student level for Number Sequences, Number Identification, Quantity Recognition, and Total Score. One hundred percent of the variance in Magnitude Comparisons was at the student level.

Effects on the Early Numeracy Distal Measure

According to Glass (1965), "Statistical significance is the least interesting thing about the results. You should describe the results in terms of

Table 2

Means and Standard Deviations for Fall and Spring Results on the TEMI-PM and TEMI-O for the Kindergarten Sample (N = 74)

	Pretest		Posttest	
Measures	*M*	*SD*	*M*	*SD*
TEMI – PM MC				
Comparison	3.67	3.16	23.33	14.46
Treatment	2.54	2.25	33.94	16.16
TEMI – PM NS				
Comparison	.83	1.24	9.38	6.86
Treatment	.62	1.07	14.53	7.33
TEMI – PM NI				
Comparison	1.83	2.26	9.79	4.04
Treatment	1.80	2.31	13.55	3.94
TEMI – PM QR				
Comparison	6.04	5.84	21.6	10.20
Treatment	5.34	5.91	30.94	10.20
TEMI – PM TS				
Comparison	12.38	9.51	64.1	28.92
Treatment	10.30	7.66	93.0	33.49
TEMI – O MPS				
Comparison	10.79	5.69	21.09	4.09
Treatment	12.76	6.65	24.23	5.09

Note. TEMI-PM = Texas Early Mathematics Inventories-Progress Monitoring; MC = *Magnitude Comparison* subtest; NS = *Number Sequence* subtest; NI = *Number Identification* subtest; QR = *Quantity Recognition* subtest; TS = Total Score; TEMI – O = Texas Early Mathematics Inventories-Outcome; MPS = *Mathematics Problem Solving* subtest.

measures of magnitude – not just, does a treatment affect people, but how much does it affect them" (cited in Kline, 2004, p. 95). Although the TEMI-O Mathematics Problem Solving group difference was not statistically significant, the Hedges' *g* effect size was .44 (β = 2.12). Further, the total variation in Mathematics Problem Solving was distributed throughout the model, with 81% at the student level and 19% at the classroom level.

Teacher Satisfaction

We explored the social validity of the interventions by asking teachers to complete a rating scale about the lessons' content and the instructional components and delivery. The average ratings were consistently high based on a 5-point scale (5 = Strongly Agree) (Median average = 4.57, ranging from 2.33 to 4.93). The statement with the lowest rating (2.33) focused on mathematics vocabulary, "The lessons need to provide more explicit instruction in mathematics vocabulary." The statement with the highest average rating (4.93) addressed the teachers' perspective about whether their students were benefiting from the lessons, "Overall, my students are benefiting from the lessons." The median average score indicated that teachers were satisfied with the lesson, suggesting that they perceived the intervention lessons as social-

Table 3
Fixed and Random Effects for Models of TEMI-PM and TEMI-O in Kindergarten Sample

	Fixed Effects		
	Predictor	Estimate	Hedges' *g* (g^*)
TEMI – PM MC	Intercept	22.818 (3.17)[a]	
	Pretest	.677 (.72)	
	Treatment[b]	11.385 (3.927)*	.73
TEMI – PM NS	Intercept	9.197 (1.58)	
	Pretest	1.079 (.74)	
	Treatment	5.457 (1.965)*	.76
TEMI – PM NI	Intercept	9.742 (.821)	
	Pretest	.418 (.197)	
	Treatment	3.772 (1.016)*	.95
TEMI – PM QR	Intercept	21.154 (2.083)	
	Pretest	.609 (.199)	
	Treatment	9.765 (2.583)*	.96
TEMI – O MPS	Intercept	21.632 (1.069)	
	Pretest	.457 (.070)	
	Treatment	2.123 (1.346)	.44
TEMI – PM TS	Intercept	61.34 (6.55)	
	Pretest	1.478 (.428)	
	Treatment	31.968 (8.137)*	.99
	Random Effects		
		Variance	Percent of Total Variation
TEMI – PM MC	Level 1 (individual)	262.031	100%
	Level 2 (class)	.000	0%
TEMI – PM NS	Level 1 (individual)	45.09	80.35%
	Level 2 (class)	11.02	19.65%
TEMI – PM NI	Level 1 (individual)	15.54	84.22%
	Level 2 (class)	2.91	15.78%
TEMI – PM QR	Level 1 (individual)	101.66	84.34%
	Level 2 (class)	18.87	15.66%
TEMI –PM TS	Level 1 (individual)	1026.251	86.76%
	Level 2 (class)	156.589	13.24%
TEMI – O MPS	Level 1 (individual)	19.825	81.22%
	Level 2 (class)	4.584	18.78%

Note. [a]Standard errors are in parentheses; [b]Reference group is comparison; * Statistically significant after Benjamini-Hochberg correction; g^* = Effect sizes were estimated as the ratio between the model-derived treatment coefficients and the unadjusted pooled within-group standard deviation across conditions at posttest. TEMI-PM = Texas Early Mathematics Inventories-Progress Monitoring; MC = *Magnitude Comparison* subtest; NS = *Number Sequence* subtest; QR = *Quality Recognition* subtest; TEMI – O = Texas Early Mathematics Inventories-Outcome; MPS = *Mathematics Problem Solving* subtest.

ly valid for their intended purposes (i.e., improving student mathematics performance in early numeracy concepts and skills). Survey items, along with the teachers' average ratings and standard deviations, are reported in Table 4.

Discussion

Successful mathematics performance in the area of early numeracy concepts and skills is fundamental for more advanced mathematics instruction at the elementary and secondary levels. However, some kindergarten students demonstrate difficulty learning the foundational concepts and skills, and thus could benefit from interventions that supplement core instruction. Such practice is particularly imperative because low mathematics performance at the end of kindergarten has been shown to predict the continuation of low performance into the elementary grades (Jordan et al., 2009; Morgan et al., 2009).

Table 4
Results of Teacher Social Validity Satisfaction Survey

Survey Statement	*M*	*SD*
The lessons' mathematical content is aligned to the state standards: number/operation; patterns, relationships and algebraic thinking; problem solving.	4.40	1.06
The scope and sequence of the lessons for each skill are appropriate for my grade.	4.73	.46
The instructional content (warm-up, 1 lesson for K, 2 lessons for first/second grades) is an appropriate amount of content for the students to learn during the designated lessons: not too much for the allotted time.	4.47	.64
Student understanding of the skills being taught in each lesson is enhanced by the use of concrete (e.g., cubes, rods) and pictorial (e.g., number line, 100s chart) materials to represent concepts.	4.80	.56
There are appropriate amounts of practice with concrete, pictorial, and abstract representations.	4.73	.46
The lessons need to provide more explicit instruction in mathematics vocabulary.	2.33	.82
Warm-ups provide helpful practice on skills.	4.40	.51
The instructions and "teacher talk" help me teach the lessons effectively.	4.67	.49
Modeling/Modeled Practice: The modeling clearly demonstrates the skill or task.	4.80	.41
Guided Practice (GP): The instruction provides the student with sufficient and appropriate practice.	4.53	.64
Independent Practice (IP): The lessons provide opportunity for student to perform the activity independently so that I can evaluate student learning.	4.80	.41
The error correction procedures provide helpful ideas for correcting errors.	4.60	.63
There is sufficient time allotted for each part of the daily lesson.	4.27	.88
The materials for the lessons are manageable (not too much to change across the lessons and thus eat up time).	4.07	1.03
The Independent Practice progress monitoring activity provides sufficient information to help me decide if the student learned the instructional content presented.	4.47	.83
The daily check sheet is a useful tool to help me "at a glance" determine how individual students are progressing daily.	3.93	1.03
The Aim Checks provide helpful information about student progress with instruction.	4.80	.41
Overall, my students are benefiting from the lessons.	4.93	.26

This study examined the effects of Tier 2 interventions on early numeracy concepts and skills delivered by classroom teachers in kindergarten to students who were identified as having mathematics difficulties and teacher satisfaction about the intervention.

The interventions consisted of instructional design and delivery features found to produce positive academic outcomes for struggling students (e.g., Swanson et al., 1999) along with practices that contribute to mathematics performance (e.g., visual representations, mathematics vocabulary instruction).

Effects of the Intervention on Early Numeracy Performance

For the first research question, overall results on the TEMI-PM were generally favorable. That is, students in the treatment condition outperformed the comparison students ($g^* = .99$) and demonstrated statistically significantly higher scores on the TEMI-PM Total Score and all four subtests. These findings are similar to those of Clarke et al. (2016), who found an effect size of .30 and statistically significant difference between groups favoring the treatment group as measured by the EN-CBM.

The hypothesis that students in the treatment group would outperform students in the business-as-usual comparison group was confirmed. Thus, when provided intervention targeting specific early numeracy concepts and skills, young students seem to benefit from this supplemental instruction, a finding that is encouraging considering the prospect of further math struggles without intervention (Jordan et al., 2009; Morgan et al., 2009). Additionally, about 80% to 100% of the variance was accounted for at the student level. These findings are educationally significant and clinically meaningful according to What Works Clearinghouse (2014) guidelines (http://ies.ed.gov/ncee/wwc/pdf/wwc_version1_standards.pdf).

Although these scores are positive indicators for the effects of the intervention for the treatment group, we cannot ignore the potential of false positives when identifying young children with risk status; thus, we undoubtedly must attribute some of the possible effects to expected growth for typical students. Overall, the results are encouraging and show promise for the intervention. The positive effects for the Magnitude Comparisons subtest were particularly important because this aspect of mathematics may be a good indication of future understanding and performance with place value (Gersten et al., 2012). For young children who are entering formal schooling and demonstrating mathematics difficulties, critical numeracy concepts and skills paired with the use of mathematically correct vocabulary potentially can help them benefit from intervention that may reduce further mathematics difficulties. For example, findings from the National Mathematics Advisory Panel (NMAP; 2008) stressed the importance of providing early intervention that uses effective instructional practices in the primary grades, especially for at-risk students.

Effects of the Intervention on Broad Mathematics Concepts and Skills

For the second research question, the effects of the intervention on the TEMI-O MPS, a broad measure of mathematics performance, were analyzed. The findings were educationally significant ($g^* = .44$), but the group difference was not statistically significant. Although disappointing, we were not surprised by these findings, because the intervention was not focused on a broad range of mathematics concepts and skills (e.g., geometry, measurement) as measured by the TEMI-O MPS. Thus, the hypothesis of no significant differences between groups was confirmed. Students' ability to generalize their knowledge of early numeracy concepts and skills to broader mathematics measures remains an area that warrants more research to determine how students can learn generalizations that will enable them to transfer knowledge to novel measures and other mathematics topics.

Teacher Satisfaction With the Early Numeracy Intervention

In response to the third research question, teachers generally reacted favorably to the lessons. It is important that lessons such as those used in this study incorporate positive instructional features that combine to promote student growth. In the area of mathematics instruction, general education elementary classroom teachers may not have received training on the features of instructional delivery included in the lessons because more emphasis is placed on a constructivist approach to mathematics teaching in general education teacher preparation programs. Thus, it was encouraging to find that the teachers responded well to the elements of effective instruction (e.g., modeling, checking for understanding, multiple practice opportunities) that were incorporated into the lessons. The hypothesis that the treatment teachers' satisfaction, on average, would receive high ratings because of the instructional design and delivery practices that were included in the lessons was confirmed.

Limitations and Future Research

Although the study was well designed and implemented, three primary limitations need to be addressed in future studies. First, additional fidelity observation are warranted. We collected fidelity data three times; future research efforts should examine fidelity more often to ensure that the intervention is being administered as intended. Relatedly, no data were collected on inter-observer agreement with the fidelity ratings. Second, additional observations of comparison group instruction should take place so that comparisons can be made more precisely across interventions. Third, social validity data should be gathered from students who are participating in the intervention, in addition to the perspectives of their teachers. Their perspectives about the lessons are equally valid as their teachers' views.

Educational Implications

Findings from this study demonstrate that students who are struggling with early numeracy concepts and skills can benefit from Tier 2 intervention; thus, both the amount of instructional time applied and the intervention components hold promise for future intervention work for struggling students. Noteworthy, the classroom teachers were responsible for delivering the intervention to small groups of students rather than a mathematics interventionist who pulls students from various classrooms for intervention work. The classroom teachers in this study received intensive professional development at the onset of the intervention program and coaching during the school year (albeit on a limited basis) from project staff. The overall findings suggest that teachers can be effective interventionists even in the early stages of adopting and implementing a new intervention for at-risk students when the program is part of their daily mathematics routine across the school year.

References

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

Baroody, A. (1990). How and when should place value concepts and skills be taught? *Journal for Research in Mathematics Education, 21*(4), 281-286.

Benjamini, Y., & Hochberg, Y. (1995). Comparison ling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series* B, 57, 289-300.

Benoit, L., Lehalle, H., & Jouen, F. (2004). Do young children acquire number words through subitizing or counting? *Cognitive Development, 19*(3), 291-307.

Bryant, B. R., Bryant, D. P., Porterfield, J., Dennis, M. S., Falcomata, T., Valentine, C., Brewer, C., & Bell, K. (2014). The effects of a tier 3 intervention for second grade students with serious mathematics difficulties. *Journal of Learning Disabilities, 49*(2),176-188. doi:10.1177/0022219414538516

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

Bryant, D. P., Roberts, G., Bryant, B. R., & DiAndreth-Elkins, L. (2011). Tier 2 early numeracy number sense interventions for kindergarten and first-grade students with mathematics difficulties. In R. Gersten & B. Newman-Gonchar (Eds.), *RtI mathematics (*pp. 65-83). Baltimore, MD: Brookes Publishing Co.

Clarke, B., Doabler, C. T., Smolkowski, K., Baker, S. K., Fien, H., & Strand Cary, M. (2011). *Examining the efficacy of a tier 2 kindergarten intervention report* (Technical Report 1102). Eugene, OR: University of Oregon.

Clarke, B., Doabler, C. T., Smolkowski, K., Baker, S. K., Fien, H., & Strand Cary, M. (2016). Examining the efficacy of a Tier 2 kindergarten mathematics intervention. *Journal of Learning Disabilities, 49*(2), 152-165. doi: 10.1177/0022219414538514

Clarke, B., & Shinn, M. R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review, 33*(2), 234-238.

Clements, D. H., & Sarama, J. (Eds.). (2004). *Engaging young children in mathematics standards for early childhood education.* Mahwah, NJ: Lawrence Erlbaum Associates.

Connolly, A. J. (2008). *KeyMath-3*. San Antonio, TX: Pearson.

Cooper, J. O., Heron, T. E., & Heward W. L. (2007). *Applied behavior analysis* (2^{nd} ed.). Columbus, OH: Merrill.

Dyson, N., Jordan, N.C., Beliakoff, A., & Hasinger-Das, B. (2015). A kindergarten number sense intervention for low-achieving children with contrasting practice conditions. *Journal for Research in Mathematics Education*, *46*(3), 280-319.

Fuchs, L. S., & Fuchs, D. (2001). Principles for the prevention and intervention of mathematics difficulties. *Learning Disabilities Research & Practice, 16,* 85-95. doi:10.1111/0938-8982.00010

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: U.S. Department of Education, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/

Gersten, R., Chard, D., Jayanthi, M., Baker, S., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research, 79*(3), 1202-1242. doi:10.1177/00222194050380040301

Gersten, R., Clarke, B., Jordan, N. C., Newman-Gonchar, R., Haymond, K., & Wilkins, C. (2012). Universal screening in mathematics for the primary grades: Beginnings of a research base. *Exceptional Children, 78*(4), 423-445.

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and intervention for students with mathematics difficulties. *Journal of Learning Disabilities, 38*(4), 293-304. doi:10.1177/00222194050380040301

Ginsburg, H., & Baroody, A. (2007). *Test of early mathematics ability* (3rd ed.). Austin, TX: PRO-ED.

Greenes, C., Ginsburg, H. P., & Balfanz, R. (2004). Big math for little kids*. Early Childhood Research Quarterly*, *19*(1), 159-166. doi:10.1016/j.ecresq.2004.01.010

Griffin, S. (2004). Teaching number sense. *Educational Leadership, 61*(5), 39-42.

Harcourt Assessment. (2003). *Stanford achievement test* (10^{th} ed.). San Antonio, TX: Author.

Hiebert, J., & Wearne, D. (1992). Links between teaching and learning place value with understanding in first grade. *Journal for Research in Mathematics Education, 23*(2) 98-122.

Jordan, N. C., Glutting, J., Dyson, N., Hassinger-Das, B., & Irwin, C. (2012). Building kindergarteners' number sense: A randomized controlled study. *Journal of Educational Psychology, 104*(3), 647-660. doi:10.1037/a0029018

Jordan, N. C., Glutting, J., Ramineni, C., & Watkins, M. W. (2010). Validating a number sense screening tool for use in kindergarten and first grade: Prediction of mathematics proficiency in third grade. *School Psychology Review*, *39,* 181-195.

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental Psychology, 45*, 850−867.

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology, 62,* 498-525.

Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research.* Washington, DC: American Psychological Association.

Minitab. (n.d.). *Technical support document receiver operator characteristic (ROC) curve macro.* Retrieved from http://www.minitab.com/support/documentation/answers/ROCBLR.pdf

Morgan, P. L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities, 42*, 306-321. doi:10.1177/0022219408331037

National Center for Education Statistics. (2004). *Early childhood longitudinal study, kindergarten class of 1998-1999(ECLS-K): User's manual for the ECLS-K third grade public-use data file and electronic code book (NCES 2006-001).* Washington, DC: U.S. Department of Education, Institute of Education Sciences.

National Center for Education Statistics. (2015). *NAEP 2015 mathematics: Report card for the nation and the states.* Washington, DC: U.S. Department of Education, Institute of Education Sciences.

National Council of Teachers of Mathematics. (2006). *Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for coherence.* Reston, VA: Author.

National Governors Association Center for Best Practices and Council of Chief State School Officers (NGA Center and CCSSO). (2010). *Common core state standards for mathematics*. *Common core state standards (college- and career-readiness standards and K-12 standards in English language arts and math)*. Washington, DC: Authors. http://www.corestandards.org.

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/index.html

Rasbash, J., Steele, F., Browne, W. J., & Goldstein, H. (2004). *A user's guide to MLwIN*. Bristol, UK: University of Bristol, Centre for Multilevel Modeling.

Sarama, J., & Clements, D. H. (2009). Focal points – Grades 1 and 2. *Teaching Children Mathematics, 14*, 396-401.

Stinson, D. W. (2004). Mathematics as "gate-keeper" (?): Three theoretical perspectives that aim toward empowering all children with a key to the gate. *The Mathematics Educator*, *14*(1), 8-18.

Swanson, H. L., Hoskyn, M., & Lee, C. (1999). *Interventions for students with learning disabilities: A meta-analysis of treatment outcomes.* New York, NY: Guilford Press.

Texas Education Agency/University of Texas System. (2008a). *Texas early mathematics inventories – Progress monitoring*. Austin, TX: Author.

Texas Education Agency/University of Texas System. (2008b). *Texas early mathematics inventories – Outcome*. Austin, TX: Author.

Texas Education Agency/University of Texas System. (2008c). *Texas early mathematics inventories – Technical manual*. Austin, TX: Author.

What Works Clearinghouse. (2014). *Procedures and standards handbook.* Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Wilson, A. J., Revkin, S. K., Cohen, D., Cohen, L., & Dehaene, S. (2006). An open trial assessment of "The Number Race," an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions, 2*(20), 1-16. doi:10.1186/1744-9081-2-20. Retrieved from http://www.biomedcentral.com/content/pdf/1744-9081-2-20.pdf

Wolf, M. M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11,* 203-214. doi:10.1901/jaba.1978.11-203

Woodcock, R. W., McGrew, K. S., Schrank, F. A., & Mather, N. (2007). *Woodcock-Johnson III normative update*. Rolling Meadows, IL: Riverside Publishing.

Wu, H. (2001). How to prepare students for algebra. *American Educator, 25*(2), 10-17.