

Running head: EVIDENCE-BASED TEACHING PRACTICES IN MATHEMATICS

Full Reference: Doabler, C. T., Nelson, N. J., Kennedy, P. C., Stoolmiller, M., Fien, H., Clarke, B., Gearin, B., Smolkowski, K., & Baker, S. K. (2018). Investigating the longitudinal effects of a core mathematics program on evidence-based teaching practices in mathematics. *Learning Disability Quarterly*, 41(3), 144–158. <https://doi.org/10.1177/0731948718756040>

Publication Date: Article first published online: February 19, 2018
Issue published: August 2018

Investigating the Longitudinal Effects of a Core Mathematics Program on Evidence-Based
Teaching Practices in Mathematics

Christian T. Doabler
Meadows Center for Preventing Educational
Risk
The University of Texas at Austin

Nancy J. Nelson
Center on Teaching and Learning
University of Oregon

Patrick Kennedy
Center on Teaching and Learning
University of Oregon

Mike Stoolmiller
Michigan State University

Hank Fien

Center on Teaching and Learning
University of Oregon

Ben Clarke
Center on Teaching and Learning
University of Oregon

Keith Smolkowski
Oregon Research Institute

Brian Gearin
Center on Teaching and Learning
University of Oregon

Scott K. Baker
Southern Methodist University

Funding:

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education through Grants R305A150037 and R305A080699, and by the National Science Foundation through Grant 1503161 awarded to the Center on Teaching and Learning at the University of Oregon.

Abstract

Accumulating research has established explicit mathematics instruction as an evidence-based teaching practice. This study utilized observation data from a multi-year efficacy trial to examine the longitudinal effect of a core kindergarten mathematics program on the use of explicit mathematics instruction among two distinct groups of teachers: one group that used standard practices in Year-1 of the efficacy trial and the core program in Year-2, and a second group that used the core program in both years. Targeted teaching practices consisted of teacher models, student practice opportunities, and teacher-provided academic feedback. Implementation of the program in Year-2 was found to increase the mean rates of teaching practices of teachers who used standard teaching practices in Year-1. Effect sizes are also suggestive of a positive impact of a second year of implementation with the core program. Implications for designing explicit mathematics programs and investigating evidence-based practices in future research are discussed.

Keywords: Explicit instruction, evidence-based practices, mathematics learning disabilities, core mathematics instruction, instructional design, direct observation

Investigating the Longitudinal Effects of a Core Mathematics Program on Evidence-Based Teaching Practices in Mathematics

In the medical field, evidence-based practices are identified through peer-reviewed research, especially well-designed clinical trials (Barratt, 2008). Medical care providers rely on the evidence generated by such studies to make effective decisions about the care of individual patients. Take for example a patient diagnosed with strep throat, which is a potentially life-threatening infection if left untreated. Following an evidence-based approach, a physician or nurse practitioner would treat the infection with an antibiotic. Assuming a typical response to treatment, the antibiotic would cure the patient within a week's time.

Establishing Evidence-Based Practices in the Field of Education

While the field of education lags behind the medical field in terms of the number of evidence-based practices approaches in circulation, major education reforms, such as the No Child Left Behind Act and the Individuals with Disabilities Education Act, have significantly increased the identification and use of evidence-based programs and practices in U.S. schools. The more recent signing of the Every Student Succeeds Act (ESSA, 2015) demonstrates a continued commitment among policymakers for the use of evidence-based interventions and instructional practices to support school improvement. In fact, the term “evidence-based” is cited 61 times in the new law. Although much remains to be learned, research conducted under the auspices of these federal initiatives have yielded practical benefits for effectively teaching America's disadvantaged children, including students with or at risk for mathematics learning disabilities (MLD).

For example, the advent of the Institute of Education Sciences (IES) became a primary impetus behind the initial efforts for establishing “what works” in the field of education. IES, the

evaluation arm of the U.S. Department of Education, conducts independent evaluations of completed research studies that target various programs, products, and practices. Specifically, it reviews existing research to determine if it meets rigorous research design standards. Findings of studies deemed to be of high methodological quality are then disseminated online via IES' What Works Clearinghouse.

In the field of special education, determining evidence-based practices can largely be attributed to the cogent work of the Division for Research of the Council for Exceptional Children (Odom et al., 2005). The Division, led by Odom and colleagues, charged a group of researchers to generate a set of quality indicators for four different research designs, including group and quasi experimental, single case, correlational, and qualitative research. For example, the proposed quality indicators for group- and quasi-experimental research designs took into consideration the extent to which studies described features such as their participants, intervention implementation and the comparison condition, technical adequacy of the outcome measures, appropriateness of the data analysis, effect size calculations, differential attrition, and follow-up effects (Gersten et al., 2005).

Explicit Mathematics Instruction as an Evidence-Based Practice for Students with MLD

Since the inception of IES and the quality indicators, educational researchers have begun to establish evidence-based programs and practices through methodologically rigorous studies (Cook, Tankersley, & Landrum, 2009). One instructional practice with a strong evidentiary basis for improving the outcomes of students with or at risk for MLD is *explicit mathematics instruction*. Explicit mathematics instruction is defined as a structured pedagogical approach that incorporates empirically-validated principles of instruction to effectively and efficiently teach

critical mathematics concepts and skills to mastery (Coyne, Kame'enui, & Carnine, 2011; Gersten et al., 2009).

Accumulating research suggests the beneficial effects of explicit mathematics instruction on the mathematics achievement of students with or at risk for MLD (Clarke et al., 2015; Dennis et al., 2016; Gersten et al., 2009). Much of the empirical research on explicit mathematics instruction involving students with MLD has been synthesized in two recent meta-analyses. We draw on this literature to briefly summarize three evidence-based teaching practices targeted in the current study.

In 2009, Gersten et al. analyzed 41 studies targeting students with MLD. Interventions were coded on seven dimensions including (a) explicit instructional techniques, (b) the use of visual representations of quantitative relations, (c) student verbalization of mathematics concepts and strategies for solving problems, (d) attention to the range and sequence of examples used during instruction, (e) frequent assessment feedback to teachers and students, (f) peer-assisted instruction, and (g) use of heuristics. Gersten and colleagues reported that the magnitude of the effect for explicit instruction was large (Hedges' $g = 1.22$, 95% CI (0.78, 1.67)).

More recently, Dennis et al. (2016) conducted a meta-analysis of group- and quasi-experimental studies between 2000 and 2014 on interventions for students with or at risk for MLD. A total of 25 interventions studies were analyzed based on five instructional approaches: (a) providing data and feedback on students' mathematics performance to teachers, (b) peer assisted learning, (c) providing students with data and feedback on their mathematics performance, (d) explicit or contextualized teacher facilitated instruction, and (e) computer-assisted instruction. Of the 25 studies, 18 included interventions that incorporated an explicit

mathematics instructional approach. Results suggested a large effect for the explicit interventions (Hedges' $g = .76$, 95% CI (0.45, 0.94).

Critical Teaching Practices of Explicit Mathematics Instruction

At its core, explicit mathematics instruction comprises three evidence-based teaching practices. First, teachers present new mathematical concepts, procedures, and skills to students through overt demonstrations and explanations. This allows teachers to convey higher-order thinking by making new and complex mathematics content conspicuous to students. Research suggests that direct modeling is an effective way to demonstrate what students are expected to do in a mathematical activity or task (Alfieri, Brooks, Aldrich, & Tenebaum, 2011; Dennis et al., 2016).

Following the teacher demonstrating a mathematical concept or skill, students then practice that objective through guided support from the teacher. Such practice opportunities involve students completing written exercises, manipulating visual representations of mathematics, and verbalizing their mathematical understanding. As students develop an understanding of the targeted content, the teacher's support is systematically withdrawn to increase students' opportunities to independently and actively practice with mathematics. A growing body of evidence suggests that student practice opportunities are essential for improving important mathematics outcomes (Doabler et al., 2015; Gersten et al., 2009). Though some of the particulars about practice are debated, it is generally agreed that practicing can build conceptual understanding and promote fluency thereby increasing the odds that students can develop long-term retention of important mathematical topics.

During explicit mathematics instruction, teachers complement student practice opportunities with academic feedback. Academic feedback is comprised of teachers providing

informational feedback to students on their performance with solving mathematical problems. When timely and specifically delivered, this evidence-based teaching practice can extend learning opportunities and help students circumnavigate known pitfalls or misconceptions (Coyne et al., 2011; Gersten et al., 2009). For example, when a student incorrectly verbalizes a mathematical answer, the teacher will immediately provide feedback to correct the error and then provide additional practice on the types of items that proved difficult for the student.

Instructional Design and its Role in Evidence-based Teaching Practices

Central to the effectiveness of mathematics instruction is the manner in which it is designed. In mathematics, *instructional design* refers to the judicious integration of critical mathematics content and empirically-validated principles of instruction (Coyne et al., 2011). When instructional designers purposefully engineer mathematics programs, they provide teachers with instructional tools that support student learning through the delivery of evidence-based practices, such as explicit mathematics instruction.

Only one study to our knowledge has investigated the impact of a core mathematics program on teachers' use of explicit mathematics instruction in kindergarten classrooms. This study examined treatment effects of the *Early Learning in Mathematics* (ELM) program during a randomized controlled trial (Doabler et al., 2014). ELM is an empirically-validated, core kindergarten mathematics program intended for use in general education settings (Clarke et al., 2015). ELM is a scripted program that centers on an explicit and systematic instructional design framework. Based on this framework, the program offers specific guidelines for how teachers can (a) overtly demonstrate and explain critical mathematics concepts and skills, (b) systematically facilitate meaningful opportunities for students to practice with targeted

mathematics content, and (3) offer timely, informational feedback to address student errors and potential misconceptions, and affirm students' correct responses.

In the ELM efficacy trial, Doabler and colleagues (2014) analyzed approximately 400 classroom observations from 129 kindergarten classrooms that were randomly assigned to treatment ($n = 68$) or control ($n = 61$) conditions. Classrooms assigned to the treatment condition implemented the ELM program, while control classrooms continued to provide standard district core mathematics instruction. Within the 129 classrooms were 2,598 kindergarten students, of whom 50% were considered at risk for MLD at the start of the school year. Results suggested that ELM classrooms provided statistically significantly higher rates of explicit instructional practices than control classrooms. Specifically, ELM teachers delivered more frequent practice opportunities for individuals and groups of students (Doabler et al., 2014).

While preliminary, the findings of Doabler et al. (2014) demonstrated initial promise of the ELM program to increase teachers' use of evidence-based teaching practices during core mathematics instruction. Yet, it is unclear from this earlier work whether the statistically significant differences found between the treatment (i.e., ELM) and control (i.e., standard district mathematics instruction) classrooms were based on ELM teachers implementing the program according to its design or whether ELM teachers simply utilized explicit teaching practices acquired in their prior teaching experience. Thus, it is important to investigate whether the ELM program improves teachers' use of evidence-based teaching practices (i.e., overt teacher models, student practice opportunities, and academic feedback) relative to their prior implementation of standard district programs and practices. A clearer understanding of how and when mathematics programs like ELM impact teaching practice can promote stronger professional development opportunities for teachers who work with students with or at risk for MLD. For instance, if

additional teaching experience with explicit mathematics programs increases the odds of teachers' uptake of evidence-based practices, researchers and curriculum developers may wish to take steps to increase the amount of professional development opportunities that teachers receive with such programs to help sustain implementation.

Purpose of the Study

As evidence-based practices continue to be studied by researchers and adopted by schools, there remains a critical need to investigate how they are actually implemented in educational contexts (Forman et al., 2013). It is important to monitor and understand how evidence-based practices are implemented because if schools and teachers only nominally adopt an evidence-based practice, but do not implement it with the fidelity, there is little reason to think the practice will affect student learning. A meta-analysis of 500 studies by Durlak and DuPre (2008), which found that the level of implementation was related to study outcome, underscores this point.

The American Psychological Association Division 16 Working Group on Translating Science to Practice recently outlined a broad research agenda to promote an implementation science in educational contexts (Forman et al. 2013). Of the eight areas the group recommended for study, two have particular relevance to the current study. These recommendations include: (1) determining the core components of existing evidence-based practices and (2) investigating the client and context variables that determine the success of evidence-based practice implementation. This study aims to examine some of the client and context variables that promote teachers' use of evidence-based practices during core mathematics instruction. Specifically, we investigate the longitudinal effects of the ELM kindergarten mathematics program (Clarke et al., 2015) on teachers' delivery of explicit mathematics instruction.

Therefore, the purpose of this study was twofold. First, the current analyses extend the work of Doabler et al. (2014) by estimating longitudinal change in observed rates of evidence-based teaching practices across Year-1 and Year-2 of the ELM efficacy trial for two distinct groups of kindergarten teachers. The first group, designated as CON-ELM, was represented by kindergarten teachers who provided standard district mathematics instruction (control) in Year-1 of the efficacy trial and the ELM program (treatment) in Year-2. The second group, designated as ELM-ELM, was represented by kindergarten teachers who used ELM in both years of the ELM efficacy trial. Second, the current study also used a cross-sectional comparison of the CON-ELM and ELM-ELM teachers in Year-2 to investigate the extent to which years of experience with the ELM program impacted rates of evidence-based teaching practices. Two related sets of research questions were addressed:

1. Questions regarding longitudinal change in mean rates of evidence-based teaching practices by teacher group:
 - a. For CON-ELM teachers who served as control teachers in Year-1 of the ELM efficacy trial, did mean rates of evidence-based teaching practices increase in Year-2 when these teachers used ELM, compared to Year-1 when they did not? Did any observed changes in mean rates affect the cross-year stability of those teaching practices?
 - b. For ELM-ELM teachers who implemented ELM in Year-1 and Year-2 of the ELM efficacy trial, did mean rates of evidence-based teaching practices increase in Year-2, when these teachers used ELM for a second year?
2. Questions regarding the cumulative impact of ELM teaching experience:
 - a. Did ELM-ELM teachers with two years of experience implementing ELM exhibit significantly higher rates of evidence-based teaching practices in Year-2 than CON-ELM teachers with only one year of experience with ELM?

Method

The current study examined classroom observation data collected during a multi-year, federally-funded efficacy trial that investigated the efficacy of the ELM kindergarten

mathematics program (Clarke et al., 2008). The ELM efficacy trial, which was conducted in 223 kindergarten classrooms from Oregon and Dallas, Texas, occurred in successive years, using largely the same sample of kindergarten teachers but different cohorts of students. In Year-1, classroom teachers were randomly assigned to use either the ELM program (treatment) or standard district mathematics instruction (control). In Year-2, all classroom teachers delivered the ELM program to their entire class. In Year-2 treatment classrooms only, the five lowest performing students also received the ROOTS program, a supplementary, small-group Tier 2 mathematics intervention (50 lessons) delivered by district-employed instructional assistants.

Thus, whereas the larger efficacy trial tested the impact of the ELM program on student mathematics outcomes, the present study investigated the effects of the ELM program on the use of evidence-based teaching practices among the CON-ELM and ELM-ELM teachers. CON-ELM teachers implemented standard teaching practices in Year-1 of the efficacy trial and then ELM in Year-2; whereas ELM-ELM teachers implemented the ELM program in both years. Thus, the primary unit of analysis for the present study is the classroom teacher.

Participants

Participants in Year-1 included teachers in 47 schools (37 public, 10 private) from 7 school districts in Oregon and Dallas, Texas. As shown in the top half of Table 1, schools and teachers varied in the number of years of participation. For example, 40 of the 47 schools participated in both years of the study. The seven schools that participated only in Year-1 of the study did so because they were unable to incorporate the supplementary ROOTS intervention into their schedule. Teachers who participated only in Year-1 did so either because they taught at a school that was unable to incorporate the supplementary ROOTS intervention ($n = 35$), or due to typical teacher mobility (e.g., teachers who changed grades or transferred to another school, n

= 25). In Year-1, 112 teachers provided full-day kindergarten, and 17 provided half-day kindergarten. One classroom had two teachers, each working a half-day schedule, and one teacher in a full-day program worked at a school that operated four days per week; all other kindergarten programs ran five days per week. All teachers participated for the full school year. The Year-1 sample included 17 bilingual education classes, but all mathematics instruction was provided in English. Average class size during Year-1 was 21 students ($SD = 3.7$).

In Year-2, all classrooms provided full-day kindergarten. The average class size in Year-2 was 21 students ($SD = 3.9$). Teacher demographics for both years are presented in the bottom half of Table 1.

Kindergarten Mathematics Instruction in the ELM Efficacy Trial

The multi-year, ELM efficacy trial comprised three conditions (Clarke et al., 2015): Year-1 treatment classrooms, Year-1 control classrooms, and Year-2 treatment and control classrooms. Below, we detail each of the conditions.

Year-1 Treatment Teachers. In Year-1, 68 kindergarten teachers in the treatment condition implemented the ELM program, a 120-lesson core mathematics program. Teachers were expected to adhere to ELM's scripted guidelines and deliver the program in whole class settings, 45 minutes per day, 5 days per week. ELM, a program designed to support the full range of learners, focuses on mathematics concepts and skills identified in the CCSS-M (2010). In particular, topics addressed in ELM are from the CCSS-M domains of (a) Counting and Cardinality, (b) Operations and Algebraic Thinking, (c) Number and Operations in Base Ten, (d) Measurement and Data, and (e) Geometry. ELM is grounded in validated principles of explicit mathematics instruction (Gersten et al., 2009). These principles include: (a) engaging students' prior understandings of mathematics, (b) providing vivid demonstrations and clear explanations

of mathematical concepts, (c) using visual representations of mathematical ideas to promote conceptual understanding, (d) providing opportunities for practice and review to promote mathematical proficiency, and (e) delivering timely academic feedback to confirm students' correct responses and address potential misconceptions.

Year-1 Control Teachers. In Year-1, 61 control teachers implemented standard district mathematics instruction. Mathematics instruction in the Year-1 control condition used a number of different published curricula and teacher-developed materials. The most commonly used curricula were Texas Mathematics, Harcourt, Everyday Mathematics, and teacher created lessons. Other implemented materials included Bridges in Mathematics, Progress in Mathematics, and Scott Foresman-Addison Wesley Mathematics. A primary focus of mathematics instruction in the Year-1 control classrooms was whole number concepts and skills. Mathematical content in the Year-1 control classrooms was delivered through a variety of mediums, including learning centers, small-group activities, and whole-class delivered instruction.

Year-2 Treatment and Control Teachers. In Year-2, 91 teachers across the treatment ($n = 43$) and control ($n = 48$) conditions delivered the ELM program. However, the treatment and control classrooms differed on the implementation of the ROOTS Tier 2 intervention. In treatment classrooms, at-risk students received the ROOTS intervention three days per week, instead of completing the ELM worksheet activity at the end of ELM instruction. All ROOTS instruction was delivered by district-employed instructional assistants. At-risk students in the control classrooms participated in ELM instruction, five days per week, including the worksheet activity. Preliminary models included tests of the effects of ROOTS instruction but no effects

were detected. Consequently, the ROOTS distinctions were ignored (see Method section for details).

ELM Professional Development. In both Year-1 and Year-2, ELM teachers participated in four 6-hour workshops. Workshops were distributed across the school year according to ELM's quarterly teacher manuals (30 lessons each). That is, in the summer and prior to the start of the school year, ELM teachers received the first workshop on Lessons 1 to 30. Then, prior to Lessons 31, 61, and 91, teachers received the second, third, and fourth workshops, respectively. Workshops shared procedures for implementing evidence-based teaching practices and provided ELM teachers active learning opportunities to deliver sample lessons.

Classroom Observations of Student-Teacher Interactions—Mathematics

The Classroom Observation of Student-Teacher Interactions—Mathematics (COSTI-M) is a validated, low-inference observation tool that is designed document the frequency of evidence-based teaching practices during mathematics instruction (Doabler et al., 2015; Smolkowski & Gunn, 2012). Specifically, the COSTI-M targets the occurrences of four evidence-based teaching practices: (i.e., teacher models, academic feedback, group responses, and individual responses). All four teaching practices are captured in real time and coded in a serial fashion.

Teacher models were operationalized as a teacher's explanations, verbalizations of thought processes, and physical demonstrations of mathematical content. For example, observers coded a teacher model if the teacher explicitly described the structural features of an "add to" word problem (i.e., an action element where the quantity is increased). Academic feedback was operationalized as a teacher's verbal reply or physical demonstration to affirm or clarify a

student response. For example, observers recorded an academic feedback code if the teacher restated a correct answer, such as “Yes Bailey, six plus one equals seven.”

Group response opportunities were defined as a mathematics-related verbalization produced by two or more students in unison. Examples of group responses include three students concurrently identifying the attributes of a geometric shape. Individual response opportunities were coded whenever a single student had the opportunity to verbalize or physically demonstrate her mathematical thinking, such as when a teacher asked a specific student to answer a mathematical question (e.g., “Clara, can you use the place value blocks to show 15?”). Observers also coded an individual response when the teacher posed a question to the entire group, in which it was implied that an individual student would be asked to provide an answer or response (e.g., “Who can explain how to solve the problem?”; one or more students raise hands and the teacher calls on one student to respond). To avoid coding extraneous responses not elicited by the teacher, group and individual responses had to be preceded by teacher-posed, mathematics questions or requests. In the current study, rates per minute for the four teaching practices were computed as the frequency of the behavior divided by the duration of the observation in minutes. The COSTI-M has evidence of predictive validity with a standardized mathematics achievement measure ($p = .004$, Pseudo- $R^2 = .08$) and a battery of mathematics curriculum-based measures ($p = .017$, Pseudo- $R^2 = .05$; see Doabler et al., 2015).

Classroom Observation Procedures

Core mathematics instruction was observed in all participating classrooms in the multi-year, ELM efficacy trial (Clarke et al., 2015). In both Year-1 and Year-2, trained observers conducted classroom observations in the fall, winter, and spring of each respective school year, with approximately six weeks of instruction separating each observation round. One observation

was planned per classroom for each observation round. Across Year-1 and Year-2, a total of 658 observations were completed; 17 planned observations were missed due to scheduling conflicts or teacher absences. Observations were conducted during core mathematics instruction and observers coded for the entire instructional period. All classrooms in both Year-1 and Year-2 committed to teaching mathematics at least 45 minutes per day. The average duration of observations was 46.86 minutes in Year-1 and 37.84 minutes in Year-2. All observations were scheduled in advance and were not specific to mathematical content (e.g., geometry), lesson number, or a particular instructional day (e.g., start or end of a weekly math unit).

Observers included former educators, doctoral students, faculty members, and experienced data collectors. In each project year, observers received 14 hours of training across three sessions, including an initial training lasting 6 hours. Two 4-hour follow-up trainings were conducted prior to the winter and spring observation rounds to help minimize observer drift and increase inter-observer reliability. Training focused on kindergarten mathematics instruction and procedures associated with the use of the COSTI-M. Prior to observing independently, observers were required to complete a video reliability checkout and a real-time classroom checkout with a trained research team member. Using a smaller/larger inter-observer reliability index (Hintze, 2005), all observers met an agreement criterion of .85 or higher with both checkouts.

Inter-observer Reliability Estimates

On 137 occasions (Year-1 = 74, Year-2 = 63), two observers collected data simultaneously in the same classrooms to assess inter-observer reliability. To more rigorously measure inter-observer reliability, we calculated intra-class correlations (ICCs), which represent the proportion of variance associated with the observation occasion, opposed to the observers. In Year-1, we found ICCs of .67 for teacher models, .92 for group responses, .95 for individual

responses, and .90 for academic feedback. In Year-2, we found ICCs of .82 for teacher models, .94 for group responses, .92 for individual responses, and .82 for academic feedback. Per guidelines proposed by Landis & Koch (1977), these ICCs represent substantial to nearly perfect inter-observer reliability.

Statistical Analysis

Prior to developing our statistical models, we carefully examined the univariate distributions of the evidence-based teaching practice variables (i.e., rates of teacher models, group and individual responses, and teacher-provided academic feedback), checking for outliers and non-normal distributions. We also examined scatter plots to check bivariate distributions of the repeated measures for the same problems. Because of the modest sample size of teachers, we elected to log transform all rates to better approximate the normality assumptions underlying the latent variable models. Thus, for each rate variable, we added a small positive constant (between .25 and .50, chosen to minimize skewness) to eliminate scores of zero, took the natural log, and multiplied by 10. For brevity, we will refer to the log transformed rates as simply *the rates* except in instances where this would lead to confusion.

We evaluated our research hypotheses using a series of latent variable, 3-level models: repeated rates nested within teachers at level 1, teachers nested within districts at level 2 and districts at level 3. For each type of evidence-based teaching practice, we specified the rates from the three repeated observations nested within a study (i.e., one school year) as indicators of a single teacher level latent rate variable. All models included data from both Year-1 and Year-2, with correlations between the two teacher level latent rate variables across the years. We included these parameters because we expected the evidence-based teaching practices to correlate over time, but also to evaluate the extent to which the introduction of ELM to CON-

ELM prior to the start of Year-2 disrupted the stability of these teaching practices compared to ELM-ELM. Factor loadings were all constrained to be equal to one and the factor mean was set to zero, making the teacher level model a random intercept model. The latent variable model at the district level had the same form as the teacher level model. A path diagram is shown in Figure 1 that illustrates the basic features of the 3-level model.

Given the consistency of the ELM program's instructional design, we had no reason to expect changes across the three occasions of observation within a school year. That is, we expected most of the change in teacher behavior to occur after ELM teachers received the initial PD workshop and first began implementing ELM. Accordingly, we constrained all residual variances and indicator intercepts to be equal, although these constraints were checked for compatibility with the data. Results of preliminary models testing for potential group differences in the residual covariance structure indicated that any potential effects of ROOTS in Year-2 were not large enough to be meaningful or detectable with our sample size, so we ignored the ROOTS distinction for the remainder of the analyses. We also checked for higher order clustering effects due teachers being nested within schools within districts and found that school effects were very small and never significant given district effects.

We fit 3-level models using LMER (Bates, Maechler, Bolker & Walker, 2015) or LME (Pinheiro & Bates, 2000) in R (R Core Team, 2016) with the standard assumption of MAR for missing data and restricted information maximum likelihood estimation (REML). All p -values are two-tailed and all p -values from standard tests were replaced by p -values based on parametric bootstrapping (Bates et al., 2015; Halekoh & Højsgaard, 2014) based on 1800 bootstrap replications for each model.

We first modeled the rate of each evidence-based teaching practice using an unconstrained baseline model that was stratified by status (CON-ELM vs. ELM-ELM), which allowed separate sets of parameters in both groups and in both years within a group. We then compared the fit of the baseline model to one with complete covariance constraints, and made comparisons between groups based on the model that best fit the data.

Results

Table 2 provides sample sizes, means, standard deviations, and skewness and kurtosis statistics for each repeated rate variable (raw rate per minute scale, no log transformation) in the CON-ELM and ELM-ELM groups. For example, on the first observation occasion in Year-1, control classrooms provided an average of approximately one group response per minute (.94), whereas ELM classrooms provided approximately two group responses per minute at that same time (1.97), more than twice as many response opportunities. The rate scores from each occasion of observation did not suffer from floor effects. Across all occasions, no rate variable in either group had more than 4% of the observed values that were exactly zero. As noted above, however, rates were somewhat positively skewed and so were log transformed to better approximate normality. Figure 2 shows the mean log transformed rates by group across all six observation occasions in the ELM efficacy trial (Year-1 = occasions 1-3, Year-2 = occasions 4-6). Group differences are readily apparent in Year-1 (i.e., the ELM teachers are consistently higher than the CON teachers), as is the increase in mean logged rates for the CON group in Year-2.

Model Fit

For all rates, we found no evidence for differences in the rate level residual variance (variances of the e latent residual variables in Figure 1) either across groups or across years.

Nested chi-square tests indicated that differences in rate level residual variances across the CON-ELM and ELM-ELM groups were non-significant and the 95% confidence intervals (CI) for the rate level residual variances were highly overlapping (for the four rates, from 70 to 98% of the smaller 95% CI was overlapped by the larger 95% CI). Additional details are reported in Table 3. For group response opportunities, the assumption of a stable mean level across the occasions within a year did not hold for the ELM-ELM teachers in Year-1. The first observation occasion in Year-1 for the ELM-ELM teachers was significantly higher than occasions two and three, which is clearly apparent in Figure 2. When we modified the group response opportunities baseline model to account for this, the fit improved substantially.

The complete covariance constraints model fit no worse than the unconstrained model for all rates and hence is our preferred model on the basis of parsimony. Note that the fact that the complete covariance constraints model is the preferred model means that the ELM intervention did not significantly disrupt the Year-1 to Year-2 stability of evidence-based teaching practices. Because we had no hypotheses about district effects, we summed teacher and district variances and teacher and district covariances to get overall latent variance and overall longitudinal latent correlation to compare to rate level residual error variance. The overall latent longitudinal correlations were all significant and were .34, .72, .53 and .83 respectively for demonstrations, feedback, group responses and individual responses. The reliabilities for a single occasion of observation (intra-class correlations or ICC) are the ratio of overall variance to total variance (overall variance plus rate level residual error variance), and were .21, .37, .34 and .36 respectively for demonstrations, feedback, group responses and individual responses. These ICC's imply reliabilities for three occasions of observation that are quite modest, .44, .64, .61, and .63 respectively for demonstrations, feedback, group responses and individual responses.

Because not all teachers had three occasions of observation, these reliabilities are upper limits, which justifies the latent variable approach.

Research Question 1

Parameter estimates for each latent rate model are reported in Table 4. To test our hypotheses about longitudinal differences in evidence-based teaching practices, we conducted hypothesis tests on the mean contrasts of interest from our preferred model for each rate. Results are shown in the top three panels of Figure 3 using Hedges' g values on the log transformed scale, where mean contrasts are divided by an estimate of the population standard deviation for the log transformed latent rate. In the top panel of Figure 3, longitudinal change in the CON-ELM group was positive and significant for all four rates. For the ELM-ELM group (3rd panel from the top of Figure 3), although change was positive for all four rates, only for two rates, group responses and teacher demonstrations, was the magnitude of change big enough to be significant. The 2nd panel from the top of Figure 3 shows the contrast between longitudinal change for ELM-ELM vs CON-ELM and as is apparent, the contrast is consistently negative which indicates that CON-ELM changed significantly more for three of the four rates, teacher demonstrations being the only non-significant contrast. For academic feedback, group responses, and individual responses, longitudinal change effect sizes for the CON-ELM teachers were more than twice as large as those for the ELM-ELM teachers. The effect size differences likely reflect the immediate impact of ELM on the use of evidence-based teaching practices of teachers who delivered standard-district practices and programs during the previous year. For teacher demonstrations, the amount of growth from Year-1 to Year-2 did not differ across the ELM-ELM and CON-ELM teachers. This finding was consistent with the lack of differences in teacher models between ELM-ELM and CON-ELM teachers in Year-1.

Research Question 2

Results from the year-2 cross-sectional comparison shown in the bottom panel of Figure 3 indicate that ELM-ELM teachers had non-significantly different rates compared to the CON-ELM teachers in Year-2. However, effect sizes (Hedges' g) for teacher demonstrations, group responses, and academic feedback were moderate, 0.33, 0.42, 0.43, respectively, while the corresponding effect size for individual responses was quite small, -0.10. In summary, although the evidence is not strong, the moderate effect sizes suggest that a second year of teaching experience with the ELM program had some impact on teachers' modeling of mathematical content, facilitating group practice opportunities, and providing academic feedback. No such evidence is apparent for individual responses.

Discussion

The purpose of the current study was to examine longitudinal and cross-sectional differences of observed rates of evidence-based teaching practices for two distinct groups of teachers in the ELM efficacy trial. The first group (CON-ELM) was represented by kindergarten teachers who provided business-as-usual instruction in Year-1 of the efficacy trial and the ELM program in Year-2, while kindergarten teachers who used ELM in both Year-1 and Year-2 of the ELM efficacy trial represented the second group (ELM-ELM). Data were analyzed using a 3-level, latent variable model (rates within teachers, teachers within districts and districts).

For our first research question, we tested the longitudinal effects of ELM on the teaching practices of teachers who served as control teachers in Year-1. We hypothesized that the mean rates of teaching practices would increase with the introduction of ELM in Year-2 for the CON-ELM teachers. Our results indicate that teachers who were new to ELM in Year-2 facilitated higher levels of all four rates relative to their instructional delivery documented the previous

year. These findings suggest that ELM significantly increased the use of evidence-based teaching practices for teachers who served as control teachers in the year prior. These results are noteworthy because they lend support not only to the robustness of the reported effects in Doabler et al. (2014) but also to the rationale for teachers using explicit mathematics programs to facilitate evidence-based practices during core mathematics instruction.

We were also interested in whether the introduction of ELM in Year-2 for the CON-ELM teachers would lower the cross-year stability correlation compared to teachers who used ELM in both years. We did not find this to be true. In other words, for example, teachers who tended to provide relatively less frequent group response opportunities tended to continue doing so, even after implementing ELM. Future research with an additional control group that does not implement ELM could more rigorously test this hypothesis.

For our second research question, we tested whether the amount of experience teaching ELM (i.e., 1 year vs. 2 years) affected the mean rates of evidence-based teaching practices. Our cross-sectional hypothesis was that the number of years of experience with ELM would impact rates of teaching practices. That is, we anticipated that teachers who implemented ELM for two consecutive years would have higher mean rates in Year-2 than those teachers who provided standard district practices in Year-1 and then ELM in Year-2. While results for this research question were non-significant, interpretation of the effect sizes suggest that a second year of ELM teaching experience helped teachers facilitate higher rates of evidence-based teaching practices than teachers who were new to the program in Year-2. For example, the effect sizes for teacher demonstrations, group responses, and academic feedback were 0.33, 0.42, 0.43, respectively. Effect sizes differences between the ELM-ELM and CON-ELM teachers for individual responses were small ($g = -0.10$). Collectively, these effect sizes suggest that as

teachers gain experience teaching explicit mathematics programs, they become more fluent in demonstrating and explaining mathematical concepts, managing student practice opportunities, and offering timely academic feedback. Finally, because the current study's sample size was modest it may have impacted our capacity to detect statistically significant differences in the mean rates. Future research with larger samples of teachers is therefore warranted.

Limitations

Interpretation of our findings must be viewed in light of several limitations. First, the study used a modest sample size, including only 151 kindergarten teachers. Future efficacy research with larger sample sizes is therefore warranted. Additionally, the observation plan for Year-1 and Year-2 entailed a maximum of three observations per classroom. This decision was based on available resources in the ELM efficacy trial. Although additional observations may have provided more precise estimates of teachers' use of evidence-based teaching practices, direct observations are often resource intensive, particularly in large-scale efficacy trials implemented in different geographical regions. Future research may consider video recording classroom instruction to reduce the financial burden of real-time observations. A third limitation relates to the possibility of influencing teachers' behaviors through direct observation. However, concerns of introducing a type of "reactivity" among participants were significantly reduced in this study given that information about the kinds of data collected during the classroom observations was masked and not shared with participating teachers. Interpretation of our results for teacher models were also complicated by the somewhat lower reliability estimates in Year-1, relative to other evidence-based teaching practices, which prompted us to provide additional training for documenting teacher models in Year-2. Thus, it take more training to get observers to reliably identify a teacher model than other evidence-based practices.

Another limitation is the lack of qualitative information we report about the observed evidence-based teaching practices. Comprehensive measurement systems, as noted by Douglas (2009), document a broad range of classroom features, including instructional quality. Although the observation system used in the ELM efficacy trial measured the quality of evidence-based teaching practices, these data were not included in the current analyses because different quality measures were administered in Year-1 and Year-2. Consequently, an investigation of the longitudinal differences in the quality of evidence-based teaching practices both within and between the CON-ELM and ELM-ELM groups was not possible. Additional research using a common measure of observed instructional quality is therefore needed.

Relatedly, the COSTI-M was developed and validated as a low-inference, frequency count observation measure (Doabler et al., 2015; Smolkowski & Gunn, 2012). While the COSTI-M documents the explicit teaching practices identified in the accumulating knowledge base for effectively teaching students with or at risk for MLD (Dennis et al., 2016; Gersten et al., 2009), it does not capture the complexity and duration of a teacher's overt demonstrations and explanations. However, as noted by Snyder et al. (2006), conducting rigorous intervention efficacy trials "entails making choices about measurement strategies to adequately capture each construct or element as defined by the theories informing the intervention" (p. 44). Against that backdrop, the principal investigators of the ELM efficacy trial selected the COSTI-M because it directly maps onto the theoretical mechanisms that are hypothesized to guide the ELM program's theory of change (see Clarke et al., 2015). At the core of ELM's theory of change is teachers' frequent use of explicit instructional teaching practices. Regardless, deeper investigations are needed to further unpack the different aspects of teacher modeling, including the frequency, complexity, and duration of this evidence-based teaching practice.

Implications for Research and Practice

Overall, these results suggest that core mathematics programs with purposefully-planned instructional design frameworks can increase teachers' use of evidence-based teaching practices during core mathematics instruction. Such teaching practices offer key opportunities for students to (a) receive demonstrations about new and complex mathematics concepts, (b) verbalize and demonstrate their mathematical understanding, and (c) obtain timely feedback from teachers to correct misconceptions or affirm mathematical thinking. In light of these findings, we believe this study raises several interesting considerations for teaching students with or at risk for MLD.

First, a growing body of research suggests that explicit mathematics instruction improves the mathematics achievement of students with or at risk for MLD (Dennis et al., 2016; Gersten et al., 2009). Yet, studies also suggest that features of this evidence-based instructional approach are largely missing from many of the mathematics programs used to teach struggling learners (Bryant et al., 2008). Therefore, we encourage curriculum developers to consider embedding explicit teaching practices within mathematics programs. For example, because mathematical proficiency is mostly comprised of students' ability to express and show their mathematical thinking and understanding, curriculum developers should ensure that programs offer frequent, cognitively-demanding opportunities for students to practice (e.g., mathematics verbalizations).

Another implication from the current study pertains to teaching experience with explicit mathematics programs. We found that, with exception of individual responses, two years of experience with the ELM program increased teachers' delivery of overt teacher models, group responses, and academic feedback. This finding has implications for teachers who sustain implementation of explicit mathematics programs. Additional years of experience with such programs may serve as an optimal mechanism to enhance teachers' use of evidence-based

practices. Relatedly, we encourage researchers to investigate whether continued implementation of explicit mathematics programs produces greater student mathematics achievement.

Finally, and a challenge encountered in most observational research, is the notion of volatility or day-to-day variability of classroom instruction (Douglas, 2009; Snyder et al., 2006). Research suggests that specific classroom instructional practices can be volatile on a day-to-day basis (Doabler et al., 2014). In the current study, we observed similar signs of volatility, including modest within-year ICCs for observed evidence-based teaching practices, which is what lead us to use latent variable models to avoid the downward bias on effect sizes that goes with such volatility. Given that observations were not specific to ELM lessons but rather conducted based on availability of the observation team and classroom teachers, this volatility may have been due, at least in part, to the type of mathematical content observed. For instance, it is possible that more stable teaching practices occur when instruction focuses on mathematical procedures, such as solving addition and subtraction problems, as compared to conceptually-based topics, such as understanding the concept of zero or equal. Future research in this area is warranted.

Conclusion

Over the past several decades, special education researchers have begun to establish evidence-based practices through methodologically rigorous research studies. One instructional practice found to have a strong evidentiary basis for supporting the development of mathematics proficiency among students with or at risk for MLD is explicit mathematics instruction. The current study investigated the longitudinal effects of a core mathematics program on teachers' use of four evidence-based practices associated with explicit mathematics instruction. Findings from this study suggest that the implementation of explicit mathematics programs may increase

teachers' facilitation of overt teaching models, student practice opportunities, and academic feedback. Future research is warranted to determine whether our findings replicate with other core mathematics programs and hold up to variations in Tier 2 and Tier 3 settings.

References

- Alfieri, L., Brooks, P., Aldrich, N., & Tenebaum, H. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology*, 103, 1–18. doi:10.1037/a0021017
- Barratt, A. (2008). Evidence based medicine and shared decision making: The challenge of getting both evidence and preferences into health care. *Patient Education and Counseling*, 73, 407–412. doi:10.1037/a0021017
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.
- Bryant, B. R., Bryant, D. P., Kethley, C., Kim, S. A., Pool, C., & Seo, Y.-J. (2008). Preventing mathematics difficulties in the primary grades: The critical features of instruction in textbooks as part of the equation. *Learning Disability Quarterly*, 31, 21–35. doi:10.2307/30035523
- Clarke, B., Baker, S., Smolkowski, K., Doabler, C., Cary, M. S., & Fien, H. (2015). Investigating the efficacy of a core kindergarten mathematics curriculum to improve student mathematics learning outcomes. *Journal of Research on Educational Effectiveness*, 8, 303–324. doi:10.1080/19345747.2015.1116034
- Cook, B. G., Tankersley, M., & Landrum, T. J. (2009). Determining evidence-based practices in special education. *Exceptional Children*, 75, 365–383. doi:10.1177/001440290907500306
- Coyne, M. D., Kame'enui, E. J., & Carnine, D. W. (2011). *Effective teaching strategies that accommodate diverse learners* (4th ed.). Upper Saddle River, NJ: Prentice Hall.
- Dennis, M. S., Sharp, E., Chovanes, J., Thomas, A., Burns, R. M., Custer, B., & Park, J. (2016). *A meta-analysis of empirical research on teaching students with mathematics learning*

- difficulties. *Learning Disabilities Research & Practice*, 31, 156–168.
doi:10.1111/ldrp.12107
- Doabler, C. T., Baker, S. K., Kosty, D., Smolkowski, K., Clarke, B., Miller, S. J., & Fien, H. (2015). Examining the association between explicit mathematics instruction and student mathematics achievement. *Elementary School Journal*, 115, 303–333.
doi:10.1086/679969
- Doabler, C. T., Nelson, N. J., Kosty, D. B., Fien, H., Baker, S. K., Smolkowski, K., & Clarke, B. (2014). Examining teachers' use of evidence-based practices during core mathematics instruction. *Assessment for Effective Intervention*, 39, 99–111.
doi:10.1177/1534508413511848
- Douglas, K. (2009). Sharpening our focus in measuring classroom instruction. *Educational Researcher*, 38, 518–521. doi:10.3102/0013189X09350881
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41, 327–350.
doi:10.1007/s10464-008-9165-0
- Forman, S. G., Shapiro, E. S., Coddling, R. S., Gonzales, J. E., Reddy, L. A., Rosenfield, S. A., . . . Stoiber, K. C. (2013). Implementation science and school psychology. *School Psychology Quarterly*, 28, 77–100. doi:10.1037/spq0000019
- Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research*, 79, 1202–1242.
doi:10.3102/0034654309334431

- Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children*, 71, 149–164. doi:10.1177/001440290507100202
- Halekoh, U., & H.jsgaard, S. (2014). A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models—The R Package pbrtest. *Journal of Statistical Software*, 59, 1–32.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174. doi:10.2307/2529310
- Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-plus*. New York, NY: Springer.
- R Core Team (2016). *R: A language and environment for statistical computing (Version 3.3)* [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Available from <https://www.R-project.org/>
- Smolkowski, K., & Gunn, B. (2012). Reliability and validity of the Classroom Observations of Student–Teacher Interactions (COSTI) for kindergarten reading instruction. *Early Childhood Research Quarterly*, 27, 316–328. doi:10.1016/j.ecresq.2011.09.004
- Snyder, J., Reid, J., Stoolmiller, M., Howe, G., Brown, H., Dagne, G., & Cross, W. (2006). The role of behavior observation in measurement systems for randomized prevention trials. *Prevention Science: The Official Journal of the Society for Prevention Research*, 7, 43–56. doi:10.1007/s11121-005-0020-3