Hammill Institute on Disabilities

Assessment for Effective Intervention
2019, Vol. 44(3) 197–211

DOI: 10.1177/1534508418758364
aei.sagepub.com
SAGE

Do Components of Explicit Instruction Explain the Differential Effectiveness of a Core Mathematics Program for Kindergarten Students With Mathematics Difficulties? A Mediated Moderation Analysis

Christian T. Doabler, PhD[1], Mike Stoolmiller, PhD[2], Patrick C. Kennedy, PhD[3], Nancy J. Nelson, PhD[3], Ben Clarke, PhD[3], Brian Gearin, MEd[3], Hank Fien, PhD[3], Keith Smolkowski, PhD[4], and Scott K. Baker, PhD[3,5]

Abstract

A growing body of research suggests that the effects of core mathematics instruction on student mathematics outcomes may not be uniform across different skill levels in mathematics. This study investigated the extent to which observed components of explicit mathematics instruction explained why students' initial mathematics achievement was previously found to moderate the treatment impact of an empirically validated, core kindergarten mathematics program. Instructional components examined were as follows: (a) teacher demonstrations and explanations of mathematical concepts, (b) group and individual student practice opportunities, and (c) teacher-delivered academic feedback. Findings suggest that the rate in which teachers facilitated individual student practice opportunities during core mathematics instruction explained the program's differential effectiveness. Implications in terms of differentiating practice opportunities for at-risk learners and utilizing classroom observation data to test potential mediating variables of academic interventions are discussed.

Keywords

mathematics difficulties, explicit mathematics instruction, student practice opportunities, direct observations, mediating variables, efficacy research

Core (Tier 1) mathematics instruction plays a critical role in students' development of mathematics proficiency (Agodini & Harris, 2010). At each grade level, core mathematics instruction represents the mathematics instruction that focuses on the range of mathematical standards students are expected to learn and know. This definition recognizes that core instruction takes place in general education settings and is commonly delivered by teachers using commercially available core programs. A primary task of core mathematics instruction is to target the critical mathematics concepts and skills considered essential for understanding more advanced mathematical topics. For example, in kindergarten, core mathematics instruction focuses on helping students build a deep and lasting understanding of foundational aspects of early number sense (Gersten et al., 2009; Sood & Jitendra, 2013).

Importance of Core Mathematics Instruction in Kindergarten

At all grade levels, core mathematics instruction is essential for children's mathematical learning. This is particularly true in kindergarten, when core mathematics instruction sets the trajectory for future growth in mathematics and thus

[1]The University of Texas at Austin, TX, USA
[2]Michigan State University, Marquette, MI, USA
[3]University of Oregon, Eugene, OR, USA
[4]Oregon Research Institute, Eugene, OR, USA
[5]Southern Methodist University, Dallas, TX, USA

Corresponding Author:
Christian T. Doabler, The University of Texas at Austin, 1 University Station, D5300 SZB 408B, Austin, TX 78712, USA.
Email: cdoabler@austin.utexas.edu

supports all students, including students at risk for mathematics difficulties (MD), in acquiring mathematical proficiency (Jordan, Kaplan, Ramineni, & Locuniak, 2009). For many students, the core mathematics instruction provided in kindergarten represents their initial exposure to formal mathematics instruction. Therefore, it is responsible for not only allowing typical achieving students to learn and progress successfully, but also accelerating the learning of students who enter kindergarten at risk for early MD.

Recent research has begun to document the importance of core mathematics instruction for students at risk for MD. For example, in a recent quasi-experimental study, Sood and Jitendra (2013) explored the treatment effects of a core kindergarten number sense program in five kindergarten classrooms. The program, which embraced an explicit instructional approach, targeted building students' understanding of whole number relationships. A total of 101 kindergarten students participated in the study, of which 43 were identified as at risk for MD. Sood and Jitendra reported statistically significant treatment effects on a set of number sense measures, with effect sizes (Hedges's *g*) ranging from 0.55 to 1.14. Findings also suggested that at-risk students benefited from the number sense program to commensurate to their nonrisk peers.

Differentiated Effects of Core Mathematics Instruction by Achievement Level

Although core mathematics instruction during the kindergarten year is of significant importance, mounting evidence indicates that its effects may not be uniform across different initial skill levels in mathematics (Duncan et al., 2007; Morgan, Farkas, & Wu, 2009). For instance, Duncan et al. (2007) investigated mathematics achievement data from several nationally representative samples of kindergarten children, including the Early Childhood Longitudinal Study—Kindergarten Cohort (ECLS-K) and National Institute of Child Health and Human Development Study of Early Child Care and Youth Development (NICHD SECCYD) datasets. The analyses included a collective sample of approximately 22,000 kindergarten students. Findings from the ECLS-K and NICHD SECCYD datasets suggested that students with higher mathematics achievement outcomes at the start of kindergarten showed stronger response to core mathematics instruction at third and fifth grades, respectively.

In a more recent analysis of the ECLS-K dataset, Morgan et al. (2009) found that students entering kindergarten with lower mathematical skills at the start of the school year (defined by the authors as those students who scored below the 10th percentile on a nationally normed mathematics test) demonstrated a lower achieving response to core mathematics instruction through fifth grade. In kindergarten, this differential response to core instruction may be due, at least in part, to a lack of structured and explicit opportunities for students to build early number sense.

It is encouraging, however, that a recent line of efficacy research has begun to demonstrate that it is possible for systematically designed and explicitly delivered core mathematics instruction to increase achievement among students who were initially low performing at the start of kindergarten without adversely affecting the achievement of initially higher performing students. For example, Clarke and colleagues (2015a) conducted a randomized controlled trial (RCT) to test the efficacy of the Early Learning in Mathematics (ELM) program. ELM is a yearlong core (Tier 1) kindergarten mathematics program that incorporates an explicit and systematic instructional design framework (Coyne, Kame'enui, & Carnine, 2011). Approximately 2,600 kindergarten students participated in the study, of whom 50% were considered at risk for MD at the start of school year. Blocking on school, 129 kindergarten classrooms in Oregon and Texas were randomly assigned to either treatment or control conditions. Classrooms in the treatment condition implemented ELM, whereas control classrooms continued to use standard district practices (i.e., "business-as-usual").

Using a nested time by condition analysis to account for the nesting of students within classroom, Clarke et al. (2015) found that the effects of the ELM program were moderated by students' initial mathematics achievement. Specifically, findings suggested that the ELM program had significantly stronger impact for kindergarten students who tested below the 25th percentile on the *Test of Early Mathematics Ability–Third Edition* (TEMA-3; Ginsburg & Baroody, 2003) at the start of the school year compared with students who began the year with higher TEMA-3 pretest scores. Author and colleagues also reported that treatment students who tested above the 25th percentile on the TEMA-3 at the start of the year performed commensurately relative to their nonrisk control peers. That is, the ELM program kept these typically achieving students "on track" for developing mathematics proficiency, whereas students below the 25th percentile grew more than expected, based on normative data.

Taken together, these studies (Clarke et al., 2015; Duncan et al., 2007; Morgan et al., 2009) indicate that the impact of core mathematics instruction in kindergarten may differ by students' initial skill levels in mathematics. In light of these findings, there is an urgent need to investigate particular instructional components that serve as mediating variables and may help explain the moderating effects of core mathematics programs. Of particular relevance for the current study are the core components or active ingredients of explicit mathematics programs.

Instructional Components of Explicit Mathematics Instruction as Potential Mediators

A growing body of research suggests that both typically achieving students and students with MD significantly

benefit from instruction that is systematically designed and explicitly delivered (Agodini & Harris, 2010; Gersten et al., 2009). Explicit mathematics instruction is known for facilitating scaffolded instructional interactions between teachers and students related to critical mathematics content (Hughes, Morris, Therrien, & Benson, 2017). At the forefront of this instructional approach are three core components: (a) teacher demonstrations and explanations of mathematical concepts and skills, (b) group and individual student practice opportunities, and (c) teacher-delivered academic feedback. We focus on these particular components for two reasons. First, a relatively large body of research suggests that these components are associated with increased student mathematics achievement (Doabler et al., 2015; Clements, Agodini, & Harris, 2013; Gersten et al., 2009). Second, and perhaps most importantly, the rate at which students receive overt teacher demonstrations, group and individual practice opportunities, and academic feedback during core mathematics instruction (and thus the cumulative effect over the course of the school year) may serve as potential mediating variables. As such, these instructional components may help unpack and explain why core mathematics programs like ELM are able to both accelerate the learning of students with MD and support typically achieving students in developing mathematics proficiency. In the following section, we briefly define these three instructional components and review the empirical literature base behind them.

The first component of explicit mathematics instruction targeted in this study is *teacher demonstrations.* In explicit mathematics instruction, teachers play an active and prominent role in building students' conceptual and procedural knowledge. Leading these efforts are vivid demonstrations and clear explanations of mathematical concepts, skills, procedures, and vocabulary. Research suggests that overt teacher demonstrations are an effective way to present critical academic content (Alfieri, Brooks, Aldrich, & Tenebaum, 2011). A second component examined is *academic feedback*. Academic feedback represents teachers actively monitoring students' interpretations of mathematical tasks. When academic feedback is immediate and specific, research suggests that it is an effective method for extending learning opportunities, addressing students' errors, and helping them circumnavigate known misconceptions (Hattie & Timperley, 2007).

The third component of explicit mathematics instruction focused on in the current study is *practice opportunities* for individuals and groups of students. Essential to supporting students' development of mathematical proficiency are practice opportunities that target critical mathematics concepts and skills. For example, research suggests the beneficial impact of having students use visual representations of mathematical ideas, such as counting cubes and place value blocks. Such representations help students build an important connection between the conceptual and abstract forms of mathematics (Gersten et al., 2009). Another important form of practice is student mathematics verbalizations. Accumulating research indicates the importance of frequent mathematics verbalizations (Doabler et al., 2015; Clements et al., 2013; Gersten et al., 2009b). Well facilitated, mathematics verbalizations allow students the opportunity to convey their mathematical thinking and understanding.

During core instruction, teachers can direct practice opportunities to the classroom at large as well as to individual students. When properly orchestrated, group student practice opportunities provide multiple students the opportunity to practice in unison. For example, a teacher might have a class of 25 students use base 10 blocks to represent teen numbers. Individual student practice opportunities, particularly mathematics verbalizations, permit teachers to monitor and verify an individual student's mathematical understanding. This type of student practice also serves as an optimal mechanism for teachers to differentiate instruction based on a student's needs (Gersten et al., 2009).

In sum, the core components of explicit mathematics instruction (i.e., teacher demonstrations, group and individual practice opportunities, and academic feedback) represent potential mediating variables and thus may help researchers better understand for whom, when, and why educational interventions improve desired student outcomes. Specifically, investigating such instructional components may help the field ascertain as to why students differentially respond to educational interventions. Here, we define moderating variables as student-level factors that influence the relationship between an intervention and student outcomes. The moderating variable of particular relevance for the current study is kindergarten students' initial skill level in mathematics, which Clarke et al. (2015) found to moderate the treatment effects of the ELM program. Mediating variables, for the purpose of the current study, are defined as the components of explicit mathematics instruction (i.e., active ingredients) incorporated into the ELM program.

Purpose of the Study

The purpose of the current study was to conduct a mediated moderation analysis (Preacher, Zhang, & Zyphur, 2016), using a repeated measure, latent variable approach to investigate whether and to what extent observed components of explicit mathematics instruction (i.e., teacher demonstrations, group and individual student practice opportunities, and academic feedback), both individually and in combination, might help explain why the ELM core mathematics program has been previously found to accelerate the learning of students with MD and support typically achieving students in developing mathematics proficiency (Clarke et al., 2015). By examining the potential mediating role of these instructional components, this study was anticipated to advance the field in two ways. First, if the instructional

components are found to mediate the effects of ELM, our findings may have implications for the design of mathematics programs and professional development activities for teachers. Second, we know of no other large-scale efficacy trials that have conducted a mediated moderation analysis for understanding why an evidence-based core mathematics program operates differentially across students with varying mathematics achievement levels at the start of the school year. As such, we believe this study could provide the field with an example for applying this methodology in mathematics intervention research. Two research questions were posed:

> **Research Question 1:** To what extent do rates of instructional components (i.e., teacher demonstrations, group and individual student practice opportunities, and academic feedback) *individually* mediate the observed differential effect of ELM reported by Clarke et al. (2015)?
> **Research Question 2:** To what extent does an analytic model that *combines* these rates mediate the observed differential effect of ELM reported by Clarke et al. (2015)?

Method

The current study analyzed direct observation and student mathematics outcomes data collected during the ELM efficacy trial (Clarke et al., 2015). The ELM efficacy trial took place in kindergarten classrooms from Oregon and Texas, respectively, during the 2008–2009 and 2009–2010 school years. Each year involved a different cohort of kindergarten students. We used multilevel structural equation modeling (MSEM) to test for mediated moderation effects with the ELM data (Preacher et al., 2016), including a model with random classroom-level posttest (spring) on pretest (fall) slopes as the outcome and the effects of latent instructional component rates as proximal mediators of the effects of the ELM program. This expanded mediated moderation model was expected to provide important information about potential mediation effects of differential intervention effects via components of explicit instruction. Moreover, the modeling approach (i.e., MSEM with latent variables) was anticipated to provide a more accurate and nuanced picture of the impact of ELM on student mathematics achievement (i.e., eliminate conflation of between and within effects, and reduce bias due to the modest stability of the rate of instructional components; Preacher et al., 2016).

Schools and Kindergarten Classrooms

A total of 129 kindergarten classrooms from 46 schools (32 public, 11 private, and three charter) in Oregon and Dallas, Texas participated in the study. All private and charter schools were located in Texas. Of the 129 classrooms (64 in Oregon, 65 in Texas), 68 were randomly assigned (within school) to the treatment condition (i.e., ELM program) and 61 were randomly assigned (within school) to the control condition (i.e., standard district mathematics instruction). In Oregon, 17 classrooms provided half-day kindergarten. All other classrooms provided full-day kindergarten. One full-day kindergarten classroom in Oregon met only 4 days per week. Average class sizes were 21.3 (SD = 3.7) and 20.2 (SD = 3.7) for the ELM and control conditions, respectively.

Teachers. The 129 classrooms were taught by 130 teachers (69% White, 20% Hispanic, and 11% another ethnic group), of whom all but one held a teacher certification. One full-day classroom was taught by two halftime teachers. About 98% of the teachers were female with 7 or more years of teaching experience. About half (51%) had completed college-level coursework in algebra, and 39% held a graduate degree. All teachers participated for the duration of the study, and thus the outcomes of the study were not affected by attrition.

Students. Overall, 2,708 students participated in the study (1,475 in ELM classrooms and 1,233 in control classrooms). Over half of the 2,708 students started their kindergarten year below the 25th percentile on the TEMA-3. Student demographic data were only available for those students who attended participating public schools. In the 32 public schools, an average of 76% of the student population qualified for free or reduced-price lunch programs. Students in Oregon were American Indian (1%), Asian and Pacific Islander (5%), Black (2%), Hispanic (36%), and White (56%). In Texas, students were American Indian (<1%), Asian and Pacific Islander (<1%), Black (29%), Hispanic (69%), and White (1%).

ELM

ELM is a whole-class, core kindergarten mathematics program designed to promote students' development of mathematical proficiency in counting and cardinality, operations and algebraic thinking, number and operations in base 10, measurement and data, geometry, and mathematics vocabulary. ELM is delivered via explicit instruction over the course of 120 daily lessons, each approximately 45 min in duration. To support implementation, ELM teachers received four 6-hr professional development workshops across the school year in which they practiced delivering ELM content and received feedback on facilitating teacher demonstrations, student practice opportunities, and academic feedback. A detailed description of ELM can be found in Clarke and colleagues (2015).

Control Classrooms

Teachers in the control condition implemented standard district practices, utilizing both teacher-developed and commercially available mathematics activities and curricula, and a variety of instructional strategies and formats. The most widely used programs were *Everyday Mathematics, Houghton Mifflin, Scott Foresman*, and *Bridges in Mathematics*. Similar to ELM classrooms, core mathematics instruction in the control classrooms primarily focused on whole number concepts followed by concepts of geometry and measurement.

Measures

TEMA-3. The TEMA-3 (Ginsburg & Baroody, 2003) is a norm-referenced measure of students' early number sense. Internal consistency reliabilities of the measure exceed .92 and alternate form and test–retest reliabilities exceed .80. Concurrent validity coefficients with four widely used tests of mathematics ranged from .55 to .91. In the current study, the intraclass correlation coefficient (ICC) for standard scale scores on the pretest TEMA-3 was .26, and the average reliability across all 129 classrooms was .85. We used the TEMA-3 as our primary measure of student mathematics achievement, and the observed classroom average of the pretest TEMA-3 to represent classroom-level effects of the pretest TEMA-3 on outcomes in all the multilevel structural equation models (SEMs).

Early Numeracy—Curriculum-Based Measures (EN-CBM). EN-CBM (Clarke & Shinn, 2004) is a set of four fluency-based measures of early number sense. Measures include oral counting, number identification, quantity discrimination, and strategic counting with strings of numbers. EN-CBM has evidence of predictive validity with the TEMA-3 ($p < .05$, $r = .81$; Doabler et al., 2015). In the current study, we used the total score on the EN-CBM in the fall of kindergarten as a predictor of TEMA-3 missing data.

Classroom Observation of Student–Teacher Interactions—Mathematics (COSTI-M). The COSTI-M is a frequency-based observation measure that documents real-time occurrences of four components of explicit mathematics instruction: (a) teacher demonstrations, (b) group student practice opportunities, (c) individual student practice opportunities, and (d) teacher-provided academic feedback. Teacher demonstrations represent a teacher's overt explanations, verbalizations of thought processes, and physical demonstrations of mathematical content. Group student practice opportunities represent mathematics verbalizations from two or more students. Individual practice opportunities refer to one student verbalizing or physically demonstrating her mathematical understanding. Academic feedback refers to a teacher's verbal reply or physical demonstration to affirm or clarify a correct or incorrect student response (for further details on the COSTI-M, see Doabler et al., 2015). The COSTI-M has evidence of predictive validity with the TEMA-3 ($p = .004$, pseudo-$R^2 = .08$) and the EN-CBM ($p = .017$, pseudo-$R^2 = .05$; see Doabler et al., 2015). In the analyses reported here, stability ICCs were .40, .44, .26, and .44 for individual practice opportunities, group practice opportunities, teacher demonstrations, and academic feedback, respectively. Average stability of these rates was modest (.62, .66, .45, and .65, respectively), which contributed to our decision to use a latent variable approach.

A total of 317 observations were completed, with an average observation length of approximately 46 min (SD = 19 min). Of the 64 Oregon teachers, 60 were observed 3 times, three were observed 2 times, and one was observed once. In Texas, we obtained two observations for all 65 participating teachers. All observations were scheduled in advance with participating teachers and were not coordinated to coincide with specific mathematical content. Interobserver agreement ICCs ranged from .61 to .99, indicating that observers were able to code occurrences of the instructional components reliably per guidelines proposed by Landis and Koch (1977).

Statistical Analysis

Missing data. Missing pretest TEMA-3 data at the student level were systematically related to other variables, complicating model estimation. In Oregon, limited English proficiency (LEP) status and students' fall mathematics achievement (as assessed by EN-CBM) both significantly and strongly predicted missing pretest TEMA-3. In Texas, public schools had significant and substantially higher rates of missing pretest TEMA-3 than did charter or private schools. Consequently, we included in the model two exogenous student-level predictors of missing data: (a) LEP status (as a dummy-coded indicator), and (b) fall total score on the EN-CBM; and two teacher-level, dummy-coded indicators of state-school status (Texas public and Texas charter or private, with odds ratio [OR] as the reference group). We refer to these variables as missing data covariates.

Modeling assumptions. Prior to developing our statistical models, we examined univariate distributions of the instructional component rate (per minute) and student outcome variables, checking for outliers and nonnormal distributions. We also inspected bivariate scatterplots at both the teacher and student levels to check for substantial departures from linearity and outliers. As the number of teachers in our sample was modest ($n = 129$), and all rate variables were positively skewed, we log transformed the rates (adding a small positive constant between .25 and .50 to eliminate scores of zero) to better approximate the normality

assumptions underlying the latent variable models. For brevity, we will refer to the log-transformed rates as simply *the rates* except in instances where doing so would lead to confusion.

Instructional component latent variable models. In the teacher-level latent variable models, we specified, for each instructional component, the observed rates from the three repeated observations as indicators of a single latent rate per minute variable. All factor loadings were constrained to be equal to one, all indicator intercepts were set to zero, and the factor mean was freely estimated, making the measurement model a random intercept model in which the latent factor captures differences among teachers in the rate of explicit instructional components that were stable across the school year. We tested the fit of this single latent variable model for each instructional component separately, and modified it accordingly if the fit was poor. We also tested the fit of the model with four latent variables, one for each instructional component, to estimate correlations among the latent rates. In that model, we allowed for correlations between residual influences across rates at specific time points.

Random intercept and slope models. To model the differential effectiveness of ELM on student mathematics achievement, we specified a two-level random teacher-level posttest on pretest TEMA-3 regression model. Figure 1 shows a schematic illustration of the regression lines for the ELM and control groups, exaggerated to emphasize four key differences: (a) the regression for the ELM group has a flatter slope than the regression for the control group, (b) differences between the regressions are greatest at the low end of the pretest TEMA-3 distribution (i.e., the least skilled students in ELM gained the most, relative to those in the control condition), (c) those differences diminish as the pretest TEMA-3 increases, and (d) both regressions pass through the same point at the high end of the pretest TEMA-3 distribution (i.e., the most skilled students in ELM are not penalized relative to those in the control condition).

Figure 2 shows a set of hypothetical teacher-level regressions consistent with the pattern shown in Figure 1. On average, the slopes in the ELM classrooms are flatter than the slopes in the control classrooms, even though there is variation within each group. This represents the differential effectiveness of ELM, and we hypothesized that the difference in the average slope between the groups could be explained by average differences in the observed components of explicit mathematics instruction, as modeled by our latent instructional component rate variables. In other words, we tested the extent to which the implementation of ELM and its corresponding professional development changed teacher classroom behavior, which in turn mediated the flattening effect of the ELM intervention on the posttest on pretest TEMA-3 regression.

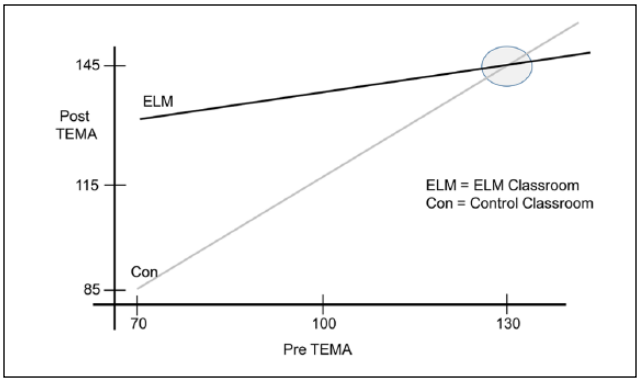

Figure 1. Schematic illustration of the differential effect of ELM, centered at a TEMA-3 pretest standard score of 130.
Note. ELM = Early Learning in Mathematics; TEMA-3 = *Test of Early Mathematics Ability–Third Edition*.

In modeling terms, our two-level model is a random intercept and slope model, but we expected the variance of the random intercept to be essentially zero when the pretest TEMA-3 score was centered at a very high score within the observed range of values but away from the extremes (i.e., 127, the 98th percentile), reflecting the fact that the intervention did not penalize initially high-skilled students. Thus, we will henceforth refer to this model as the random slope model. We estimated this model first, employing the standard assumptions that the random intercepts and slopes were multinormally distributed, as a preliminary analysis to replicate the results reported in Clarke and colleagues (2015); to verify that the estimated variance for the random slopes was significant and substantial, even after controlling for intervention status and other covariates, and check that the standard background assumptions concerning multinormality were reasonable.

To answer our first research question regarding the individual effect of each instructional component on the random slope, we estimated four separate two-level random slope models, each of which included only one of the four latent instructional component rate variables. To answer our second research question, we estimated a two-level random slope model that included three of the four latent instructional component rate variables as competing predictors of the latent slope. In this combined model, we excluded academic feedback because it was very highly correlated with group and individual student practice opportunities, due to the fact that observers were instructed to code only instances of academic feedback that followed a group or individual student practice opportunity. The full set of multilevel equations is shown in the Appendix.

All SEMs were completed with Mplus (Muthén & Muthén, 2015), using full information maximum likelihood (FIML) estimation. Where possible, we used robust maximum likelihood (MLR) estimation, which corrects the overall model fit statistic and individual parameter standard

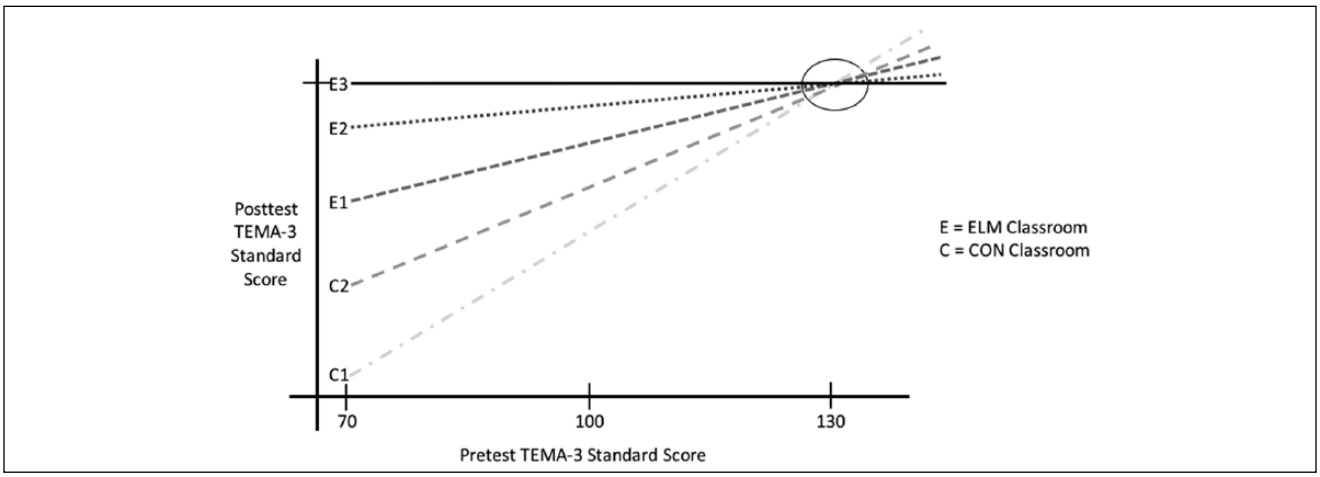

Figure 2. Hypothetical teacher-level regressions for three ELM classrooms and two control classrooms, centered at a TEMA-3 pretest standard score of 130.
Note. ELM = Early Learning in Mathematics; TEMA-3 = *Test of Early Mathematics Ability–Third Edition*.

errors for departures from multivariate normality. For models that include the pretest TEMA-3 variable in the missingness portion of the two-level model, we obtained maximum likelihood estimates using Monte Carlo numerical integration (MCNI; Asparouhov & Muthén, 2012). All *p* values are two-tailed. Model fit was evaluated using the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the chi-square *p* value.

Results

Descriptive Statistics

Our overall student sample was 2,708 students nested within 129 teachers. For these analyses, we excluded 56 students who were missing LEP status and two students who were missing all outcome data, bringing the analytic sample down to 2,650 students (98% of the overall sample). For preliminary analyses, we excluded students who were missing fall TEMA-3 data, resulting in a preliminary analytic sample of 2,212 students (82% of the overall sample). Table 1 provides descriptive statistics for the missing data covariates, student outcomes, and each instructional component by condition, and Table 2 shows the observed classroom-level correlation matrix comparing average TEMA-3 performance and instructional component rates.

Instructional Component Latent Variable Models

Results of the latent rate models for each instructional component are presented in detail in Doabler and colleagues (in press). Unique to the current analyses, we estimated a teacher-level (n = 129) latent rate model that included all four rates to obtain correlations among the latent variables. This model did not include any of the missing data covariates or intervention group status. The model fit well, χ^2 = 45.83, df = 46, p = .4796, RMSEA = 0, CFI = 1, Tucker–Lewis index (TLI) = 1.001. Correlations among the latent variables and among time-specific residual influences are shown in Table 3. Note the high correlations between academic feedback and group and individual student practice opportunities, which were .84 and .66, respectively. These correlations informed our decision to exclude academic feedback from the combined random slope model.

Preliminary Random Intercept and Slope Models

Our preliminary random intercept and slope models included intervention group status (ELM vs. control), classroom-level pretest TEMA-3, and missing data covariates, but excluded students who were missing the pretest TEMA-3 (n = 2,212). First, we fit a baseline model that included a correlated random intercept and slope, and compared it with a more restricted model with an uncorrelated random intercept and slope. The intercept-slope correlation was nonsignificant as judged by either the critical ratio in the model that included the correlation as a free parameter (z = 0.088, p = .930) or the nested chi-square test with the model that did not include the correlation (χ^2 = .0449, df = 1, p = .8322). Next, we compared the fit of the model with the uncorrelated random intercept and slope to our hypothesized model with just a random slope. The fit of our hypothesized model was almost identical (χ^2 = 0.036, df = 1, p = .8495). In contrast, eliminating the random slope but retaining the random intercept resulted in a significant degradation of fit (χ^2 = 42.55, df = 1, p < .0001). Consequently, we dropped the random intercept parameter from subsequent analyses.

Regardless of model, and consistent with results in Clarke and colleagues (2015), intervention group status had a significant flattening effect on the random slope. In other words, classroom regressions in ELM tended to be flatter than in the control group, signaling differential effectiveness (i.e., initially

Table 1. Descriptive Statistics for Missing Data Covariates, Student Outcomes, and Average Rates of Instructional Components (per Minute) by Condition.

Variable	Condition	*n*	*M*	*SD*	Skewness	Kurtosis	Minimum	Maximum
Student missing data covariates and outcomes								
EN-CBM pretest	Control	1,233	76.47	51.24	0.54	−0.49	0	230
	ELM	1,448	75.39	52.11	0.69	−0.11	0	261
LEP status	Control	1,233	0.26	0.44	1.10	−0.80	0	1
	ELM	1,448	0.30	0.46	0.87	−1.24	0	1
TEMA-3 pretest	Control	1,233	90.76	16.64	0.16	−0.21	55	144
	ELM	1,448	89.92	17.36	0.17	−0.32	55	145
TEMA-3 posttest	Control	1,233	100.48	14.93	−0.01	0.01	54	145
	ELM	1,448	102.37	14.41	−0.19	0.35	54	145
Instructional components variables								
Teacher demonstrations	Control	61	0.52	0.31	0.99	1.45	0.02	1.67
	ELM	68	0.58	0.30	0.98	0.54	0.14	1.49
Group response opportunities	Control	61	0.77	0.62	2.03	4.73	0.18	3.44
	ELM	68	1.34	0.63	0.53	−0.26	0.29	3.09
Individual response opportunities	Control	61	0.46	0.27	0.76	0.22	0.02	1.17
	ELM	68	0.65	0.38	1.07	1.28	0.15	1.88
Academic feedback	Control	61	0.50	0.32	1.34	1.80	0.08	1.55
	ELM	68	0.78	0.34	1.18	2.10	0.24	2.14

Note. EN-CBM = Early Numeracy—Curriculum-Based Measures; LEP = Limited English Proficiency; TEMA-3 = *Test of Early Mathematics Ability–Third Edition*; ELM = Early Learning in Mathematics.

low-scoring students benefited the most from ELM). Furthermore, all classroom regressions passed through the same point at a very high pretest TEMA-3 score (i.e., 127), indicating that ELM did not penalize initially high-scoring students.

Next, we attempted to replicate the same sequence of models with the full sample, using MCNI estimation. As described in Asparouhov and Muthén (2012), we encountered estimation problems when the model included a random intercept that precluded nested model testing. In some cases, models converged but the differences in log likelihoods were not positive, and in other cases, models did not converge at all. However, our hypothesized model with only a random slope consistently and quickly converged. Parameter estimates and standard errors were similar to the model that excluded students with missing pretest TEMA-3 scores, leading to the same conclusions regarding p levels and significance for all parameters. That is, excluding students who were missing the pretest TEMA-3 score did not appear to introduce noticeable bias in parameter estimates compared with models that included them, providing additional support for the use of our more parsimonious hypothesized model in subsequent analyses.

Research Question 1: Models Testing Individual Instructional Components

As with our preliminary analyses, to answer our first research question, we ran three sets of models for each latent instructional component: two models excluding students with missing pretest TEMA-3 using MLR estimation, and a third with these students included, using MCNI estimation. Each model included only one latent rate variable as a potential mediator of the ELM effect, along with the classroom level of pretest TEMA-3 and missing data covariates. Results were consistent in all three versions of our hypothesized model: Individual student practice opportunities were the only latent rate to significantly flatten the slope (est. = −0.122, $SE = 0.047$, $p < .01$). Results for group practice opportunities (est. = 0.053, $SE = 0.074$, $p > .05$), teacher demonstrations (est. = 0.114, $SE = 0.108$, $p > .05$), and academic feedback (est. = −0.075, $SE = 0.07$, $p > .05$) were not significant. Consistent with the preliminary analyses, all MCNI models that included both a random intercept and slope did not converge.

Research Question 2: Models Testing Combinations of Instructional Components

To answer our second research question, we again ran three sets of models, two excluding students with missing pretest TEMA-3 using MLR estimation, and a third including these students using MCNI estimation. A nested chi-square comparison for the models estimated on the sample after excluding students with missing pretest TEMA-3 scores revealed that our simpler hypothesized model with just a random slope fit no worse than the more general model with both a random slope and intercept (nested $\chi^2 = 0.0123$, $df = 1$, $p = .9116$), supporting the use of the more parsimonious

Table 2. Observed Classroom-Level Correlation Matrix Comparing Average TEMA-3 Performance and Instructional Component Rate Variables.

Variable	1	2	3	4	5	6
1. TEMA-3 Pretest SS	—					
2. TEMA-3 Posttest SS	.78	—				
3. Teacher demonstrations	.24	.05	—			
4. Academic feedback	.02	.03	.29	—		
5. Group student practice opportunities	.01	−.10	.49	.64	—	
6. Individual student practice opportunities	.18	.27	.08	.53	.17	—

Note. TEMA-3 = *Test of Early Mathematics Ability–Third Edition*; SS = Standard Score.

Table 3. Correlations Among Latent Instructional Component Rates (n = 129).

		Maximum Likelihood Estimation				Bayes Estimation		
Latent Rate	Correlated With	Est.	*SE*	*z*	*p*	Est.	Low 2.5 CI	High 2.5 CI
Academic feedback	Individual practice	0.663	.091	7.311	.000	0.658	0.443	0.818
Academic feedback	Group practice	0.840	.061	13.654	.000	0.832	0.692	0.924
Academic feedback	Teacher demonstrations	0.434	.151	2.874	.004	0.441	0.093	0.687
Individual practice	Group practice	0.237	.124	1.918	.055	0.251	−0.044	0.509
Individual practice	Teacher demonstrations	0.089	.186	0.476	.634	0.108	−0.293	0.456
Group practice	Teacher demonstrations	0.520	.128	4.074	.000	0.535	0.213	0.749

Note. CI = confidence interval.

hypothesized model. As before, the MCNI version of the more general model that included both a random intercept and slope generated a saddle point solution with a log likelihood that was smaller than our more restrictive hypothesized model, which precluded a nested chi-square test.

Table 4 shows the teacher-level model parameter estimates, standard errors, and p values for both versions of our hypothesized model. The results of our hypothesized model regarding the combined impact of teacher demonstrations, group practice opportunities, and individual student practice opportunities were consistent both with prior analyses and across both approaches to estimation: (a) individual student practice opportunities was the only significant predictor of the random slope; (b) the effect of ELM on the random slope was approximately half of what it was in the model that did not include teacher-level predictors and no longer significant; and (c) the indirect effect of ELM on the random slope through individual student practice opportunities was significant. Neither of the latent rates for teacher demonstrations or group student practice opportunities had significant direct or indirect effects on the random slope. Standard errors and p values were smaller in models with the larger sample using MCNI estimation.

In both models, the effect of teacher demonstrations had almost the same magnitude as individual student practice opportunities, but opposite in sign (i.e., positive rather than negative). Log transformation turns division into subtraction, that is, the log of a quotient is equal to the log of the numerator minus the log of the denominator: $\log(a/b) = \log(a)—\log(b)$. In our model, we have the random slope = $b1$ log(teacher demonstrations + .25)—$b2$ log(individual student practice opportunities + .25) + additional terms, where $b1$ is almost the same value as $b2$. This suggests that the log of the ratio of teacher demonstrations to individual student practice opportunities might be an even more powerful and parsimonious predictor of the random slope than both terms competing against each other with different effect sizes.

We tried estimating our hypothesized model with $b1$ and $b2$ constrained to be equal in magnitude but opposite in sign. The fit of this more parsimonious model was not significantly worse, and was in fact almost identical to our hypothesized model (nested $\chi^2 = 0.0394$, $df = 1$, $p = .8426$). The estimate of the constrained effect was very similar in magnitude to the unconstrained effects in our hypothesized model, but the standard error of the constrained effect shrank appreciably compared with the standard errors of both of the unconstrained teacher effects, which in turn lowered the p value to .0007. That is, it appears that constraining the effects of teacher demonstrations and individual student practice opportunities to be equal but opposite in sign may provide a more precise estimate of their combined effects on student mathematics outcomes. Although this represents a post hoc modification of the hypothesized model, meaning that the p value cannot be accepted at face value, we find the results encouraging enough to present

Table 4. Teacher-Level Model Parameters for Final Model Predicting the Slope of the Classroom-Level TEMA-3 Posttest on Pretest Regression Using Latent Rates of Instructional Components.

		Missing pretest TEMA-3 Excluded (n = 2,212)			Missing Pretest TEMA-3 Included (n = 2,650)		
Parameter 1	Parameter 2	Est.	*SE*	*p*	Est.	*SE*	*p*
Random slope regressed on	ELM	−0.025	0.038	.518	−0.025	0.035	.480
	Texas public	0.032	0.047	.488	−0.001	0.042	.979
	Group practice	−0.043	0.072	.551	−0.024	0.065	.714
	Individual practice	−0.110	0.048	.022	−0.126	0.043	.003
	Demonstrations	0.115	0.108	.287	0.098	0.101	.334
	Texas nonpublic	0.000	0.049	.993	−0.038	0.044	.394
	TEMA-3 pretest	0.050	0.033	.131	0.068	0.032	.033
TEMA-3 posttest regressed on	ELM	−0.029	0.036	.430	−0.013	0.033	.703
	Texas public	0.206	0.062	.001	0.164	0.065	.012
	Texas nonpublic	0.071	0.046	.122	0.020	0.046	.658
	TEMA-3 pretest	0.195	0.052	.000	0.235	0.058	.000
Random slope regressed on ELM via	Group practice	−0.016	0.028	.558	−0.009	0.025	.717
	Individual practice	−0.025	0.012	.039	−0.029	0.012	.014
	Demonstrations	0.011	0.013	.397	0.009	0.012	.427
Residual variances	Random slope	0.007	0.001	.000	0.005	0.001	.000
	TEMA-3 posttest	0			0		
TEMA-3 posttest	With random slope	0			0		
Group practice regressed on	ELM	0.378	0.056	.000	0.379	0.056	.000
	Texas public	−0.334	0.069	.000	−0.333	0.069	.000
	Texas nonpublic	−0.383	0.075	.000	−0.383	0.075	.000
	TEMA-3 pretest	0.016	0.075	.829	0.017	0.075	.822
Individual practice regressed on	ELM	0.229	0.065	.000	0.228	0.065	.000
	Texas public	−0.137	0.093	.138	−0.136	0.093	.142
	Texas nonpublic	0.141	0.079	.073	0.140	0.079	.076
	TEMA-3 pretest	0.030	0.083	.716	0.032	0.083	.703
Demonstrations regressed on	ELM	0.096	0.056	.085	0.095	0.056	.086
	Texas public	−0.279	0.070	.000	−0.280	0.070	.000
	Texas nonpublic	−0.207	0.082	.011	−0.208	0.081	.011
	TEMA-3 pretest	0.119	0.071	.094	0.119	0.071	.094
Individual practice with	Group practice	−0.001	0.009	.879	−0.001	0.009	.874
	Demonstrations	−0.002	0.011	.836	−0.002	0.011	.833
Demonstration with	Group practice	0.017	0.012	.153	0.017	0.012	.153
Group 1 with	Demonstrations 1	0.081	0.023	.001	0.081	0.023	.001
Group 2 with	Demonstrations 2	0.089	0.018	.000	0.089	0.018	.000
Group 3 with	Demonstrations 3	0.087	0.017	.000	0.087	0.017	.000
Individual Practice 1 with	Group practice 1	0.032	0.023	.165	0.030	0.023	.181
	Demonstrations 1	0.001	0.024	.980	−0.002	0.024	.940
Individual Practice 2 with	Group practice 2	0.055	0.016	.001	0.054	0.016	.001
	Demonstrations 2	0.041	0.022	.059	0.043	0.022	.049
Individual Practice 3 with	Group practice 3	0.006	0.018	.726	0.008	0.018	.669
	Demonstrations 3	0.034	0.020	.086	0.034	0.020	.087
Intercepts	Group practice	0.306	0.074	.000	0.306	0.074	.000
	Individual practice	−0.442	0.063	.000	−0.441	0.063	.000
	Demonstrations	−0.191	0.068	.005	−0.191	0.068	.005
	Random slope	0.660	0.047	.000	0.679	0.046	.000
	TEMA-3 posttest	6.303	0.042	.000	6.333	0.037	.000
Residual variances	Group practice	0.042	0.013	.001	0.042	0.013	.001
	Group practice 1–3	0.138	0.017	.000	0.138	0.017	.000
	Individual practice	0.074	0.020	.000	0.074	0.020	.000
	Individual Practice 1–3	0.153	0.018	.000	0.153	0.018	.000
	Demonstrations	0.032	0.016	.048	0.033	0.016	.046
	Demonstrations 1–3	0.167	0.019	.000	0.167	0.019	.000

Note. Residual variance for TEMA-3 posttest and covariance of TEMA-3 posttest with random slope are fixed to zero and shown in the table as zeros with no standard errors or *p* values to remind readers of this fact. Residual variances for group practice, individual practice, and teacher demonstrations were constrained to be equal at all three measurement occasions, and the single value is shown on the line labeled, for example, Group practice one to three. TEMA-3 = *Test of Early Mathematics Ability–Third Edition*; Random slope = random posttest on pretest TEMA-3 slope; ELM = Early Learning in Mathematics; Texas public = Texas public school indicator; Group practice = group student-practice opportunities; Individual practice = individual student-practice opportunities; Demonstrations = teacher demonstrations; Texas nonpublic = Texas charter and private school indicator; TEMA-3 posttest = random posttest on pretest TEMA-3 intercept; with = covariance between two variables.

them as an interesting possibility that should be further tested.

Discussion

Results Summary

The purpose of this study was to extend the work of Clarke et al. (2015) by examining whether rates of observed components of explicit mathematics instruction (i.e., teacher demonstrations, group and individual student practice opportunities, and teacher-provided academic feedback) would explain, both individually and in combination, the previously observed differential effectiveness of the ELM program. Thus, the current study represents the first opportunity to examine mediators (i.e., active ingredients) that underlie ELM, a high-quality, evidence-based core mathematics program (Clarke et al., 2015; Clarke et al., 2011). Below, we summarize the results by instructional component.

Individual student practice opportunities. Across a range of model parameterizations and estimation methods, we consistently found that the latent rate of individual student practice opportunities helped shed light as to why students' initial mathematics achievement was previously found to moderate the treatment impact of the ELM program (Clarke et al., 2015). When included as a teacher-level predictor, individual student practice opportunities were found to mediate the previously identified flattening effect of the ELM program on the posttest on pretest TEMA-3 regression, suggesting that students who were at risk for MD at the start of the school year benefited most when taught in ELM classrooms that offered high rates of individualized practice. In these particular ELM classrooms, teachers facilitated approximately one individual student practice opportunity per minute, representing about 45 individualized responses across each lesson. We contend that these ELM classrooms provided highly interactive learning experiences for struggling learners.

Interestingly, our findings also suggest that individual student practice opportunities were similarly important to at-risk students in some of the control classrooms. In a few cases, at-risk students in control classrooms that provided high rates of individual student practice opportunities benefited more than their at-risk peers in ELM classrooms that had low rates of individualized practice. One interpretation from this finding is that teachers in these particular control classrooms were using an explicit instructional approach to systematically structure and directly facilitate high rates of individualized practice opportunities.

However, in contrast to the core instruction delivered in most of the control classrooms, it is important to note that, in the context of ELM, individual student practice does not represent at-risk learners simply practicing sans instructional support. Rather, as ELM centers on a systematic and explicit instructional design framework (Coyne et al., 2011), it supports teachers in differentiating instruction for at-risk learners through systematic feedback loops (Raudenbush, 2008). These loops represent critical opportunities for students to receive reinforcement of their mathematical thinking and understanding. For example, after an at-risk learner in an ELM classroom independently completes a mathematical task, the teacher will provide specific, academic feedback based on the student's performance. Such feedback is intended to reinforce the practice opportunity. If the student's initial performance is incorrect, an ELM teacher will facilitate a follow-up practice opportunity that provides more instructional support. This scaffolding is intended to promote the student's development and eventual success.

In sum, an important finding of this study is that individual practice opportunities mediate the effects of a validated, core mathematics program. These results are consistent with the growing body of research on effective mathematics instruction, which suggests that individualized practice is an important predictor of mathematics achievement for at-risk learners (Gersten et al., 2009). In kindergarten, practice opportunities are at a premium because many students, particularly those from disadvantaged backgrounds, receive few opportunities to build early number sense prior to school entry (Barnes et al., 2016). Therefore, one interpretation from our results is that the implementation of explicit, core mathematics programs, such as ELM, are necessary to support the development of early mathematical proficiency among these at-risk kindergarten students. Research indicates that when such programs are systematically designed and explicitly delivered, they have the capacity to engage students with MD in structured practice opportunities related to foundational concepts and skills of mathematics (Agodini & Harris, 2010; Wang, Firmender, Power, & Byrnes, 2016).

Teacher demonstrations and group practice opportunities. With respect to the other types of instructional components, rates of teacher demonstrations did not surface as a significant predictor of the random slope. This nonsignificant finding is consistent with earlier correlational research in mathematics instruction. For example, results from Clements et al. (2013) suggest that the observed number of representations demonstrated by first-grade teachers during mathematics instruction is not significantly related to student mathematics achievement. Interestingly, the current research and the study conducted by Clements and colleagues entailed frequency count observation systems. Therefore, it may be that a more effective way to measure this component of explicit instruction is not only to document its occurrences but also its duration and complexity. Future research is needed to examine this multiprong approach.

In a similar vein, our findings indicate that group practice opportunities did not mediate the effects of the ELM program. This nonsignificant finding is also consistent with prior research on mathematics instruction (Clements et al., 2013; Doabler et al., 2015). Although group practice opportunities are an integral component of explicit mathematics instruction (Gersten et al., 2009) and were facilitated at a high rate of delivery in ELM classrooms, it is possible that the capacity of this instructional component to account for the differential effects of a core mathematics program was mitigated based on a possible threshold effect. Once these types of student practice opportunities reach a particular threshold or rate of delivery (e.g., one per min), there may be diminishing returns. In other words, the effects of group practice on student mathematics achievement may minimize above a particular rate per minute. Additional research is needed to investigate potential threshold effects.

Combination of instructional components. A tentative finding also emerged from our second research question, which investigated whether combinations of the latent rates of instructional components mediated the differential effectiveness of ELM. Results suggested that the ratio of teacher demonstrations to individual student practice opportunities may be an important medium to support the mathematics achievement of struggling students. These findings have implications for future research on the impact of student practice. For example, researchers might explore for the existence optimal ratios of individual practice opportunities to other types of instructional components, such as teacher demonstrations. If optimal ratios do exist, it may be that their effectiveness varies based on students' initial mathematics skill levels.

Implications for Practice and Research

Although preliminary, our findings share implications with previous observation research for translating direct observation results into professional development activities designed for teachers who work with students with more intensive instructional needs (e.g., Connor et al., 2009). The individual student practice opportunities coded by the COSTI-M, such as students' mathematics verbalizations and use of concrete representations of mathematical ideas, represent the types of practice opportunities teachers typically provide students during instruction. Attempts, therefore, to support teachers in effectively engaging at-risk learners in these types of practice opportunities could be a reasonable and valuable professional development objective. One such objective could be learning how to engage students with MD in cognitively challenging practice opportunities, such as mathematics verbalizations. For example, many students struggle to verbally justify mathematical answers. As such, professional development could support teachers in differentiating these individualized practice opportunities through mathematics verbalization stems or sentence starters (e.g., "Say it with me, we solved this problem by . . .").

We also believe this study has implications for researchers who utilize direct observation data to systematically test the theoretically specified components of academic interventions. Even when observers are well trained to a high degree of interobserver agreement, the temporal stability of instructional components from one occasion to the next can be low (Smolkowski & Gunn, 2012; Doabler et al., 2015; Ho & Kane, 2013). A pattern of low temporal or situational stability is a persistent characteristic of behavioral observation systems (Heyman, Lorber, Eddy, & West, 2014). In the context of characterizing time stable (across the academic year) differences in instructional rates across teachers, low temporal stability can lead to serious attenuation of estimated effect sizes.

One solution is to conduct enough observations that an average rate based on an additive composite across observations has adequate stability. The literature on frequency-based observation measures suggests at least 6 to 12 observation occasions are required to reach reliability of .80 for intraclass correlations in the range of .40 to .25 (ICCs; (Smolkowski & Gunn, 2012; Doabler et al., 2015). However, this option is not likely to be very feasible in many situations, especially given the potential assessment burden on teachers and the expense of conducting direct observations during large-scale efficacy trials. As the current study shows, a more cost-effective approach is to conduct enough observations to define a latent instructional component rate variable. This involves more analytic complexity, but may be a more realistic option for reducing or eliminating attenuation due to low stability.

Limitations

This study had several limitations that should be considered when interpreting our findings. First, only two observations were conducted in the Texas classrooms. Although the number observations in Texas were limited to accommodate other research priorities of the ELM efficacy trial, more observation data may have provided the instructional components greater explanatory power. Second, although we investigated a robust set of observation data, the current study did not include data on instructional quality. Although these types of data are highly important (Pianta & Hamre, 2009), they were not investigated in this study for two reasons. Most importantly, our research hypotheses concentrated specifically on the frequency of explicit teacher demonstrations, teacher-provided academic feedback, and group and individual student practice opportunities. This decision was based on the strong body of evidence that supports the regular use of explicit mathematics instruction

when teaching students with MD (Gersten et al., 2009). Our second reason was based on the fact that the ELM efficacy trial elected to administer different observation measures of instructional quality in the Oregon and Texas classrooms. One measure captured global aspects of instructional quality such as classroom management and the learning environment, whereas the other measure had a much narrower focus, documenting the quality of instructional features such as the pace of instruction, transitions between activities, and student engagement. The differing priorities of the two quality measures thus precluded us from incorporating these data in the current analyses.

A third limitation was missing pretest data. Across conditions, approximately 16% of students with posttest data were missing data on the TEMA-3 pretest. To help minimize the potential for bias, we incorporated several missing data covariates (e.g., LEP status). However, these variables were also missing for some students. Thus, although we made substantial efforts to minimize the impact of missing data on our analyses, a more complete student dataset may have helped better explain the differential effectiveness of the ELM program.

Conclusion

Investigating the interplay between moderating and mediating variables can provide researchers with critical information about the "black boxes" of educational interventions (Rothman, 2013). Specifically, it can help the field ascertain as to why an educational program does or does not lead to desired student outcomes among particular subgroups of students. In efficacy trials, where treatment effects fail to emerge, researchers can use their understanding of such mediators to refine or revise mathematics programs. When mathematics programs like ELM facilitate targeted outcomes, evidence on mediating variables, such as overt teacher demonstrations, student practice opportunities, and academic feedback, can lead to ways to make core mathematics instruction more effective for the full range of learners.

Appendix

Multilevel Structural Equation Model (SEM) Equations

Structural equations for the most general version of our multilevel SEM are shown below, although limited to the inclusion of a single rate variable for simplicity. The variables are subscripted with *i* for students, *j* for teachers, and *k* for occasions of observation. Note that although repeated measures for students and occasions of observation are both nested under teachers, they are distinctly different variables and must be subscripted differently in multilevel SEM. *T1*, $\overline{T1}$, *T2, L, C* and *I* designate, respectively, the pretest TEMA-3, the classroom average pretest TEMA-3, the posttest TEMA-3, LEP status, the CBM math score, and an observed rate of teacher–student interaction. Note that *T1* is centered about the 98th percentile score of the T1 distribution. See the methods section for the rational for this centering constant.

$$T2_{i,j} = \beta_{0,j} + \beta_{1,j}\left(T1_{i,j} - 127\right) + \beta_3 L_{i,j} + \beta_4 C_{i,j} + r_{i,j}$$

$$I_{k,j} = \beta_{2,j} + e_{k,j}$$

$$\beta_{0,j} = \gamma_{0,0} + \gamma_{0,1}\beta_{2,j} + \gamma_{0,2}\overline{T1}_j + \gamma_{0,3}P_j + \gamma_{0,4}N_j + \gamma_{0,5}E_j + u_{0,j}$$

$$\beta_{1,j} = \gamma_{1,0} + \gamma_{1,1}\beta_{2,j} + \gamma_{1,2}\overline{T1}_j + \gamma_{1,3}P_j + \gamma_{1,4}N_j + \gamma_{1,5}E_j + u_{1,j}$$

$$\beta_{2,j} = \gamma_{2,0} + \gamma_{2,2}\overline{T1}_j + \gamma_{2,3}P_j + \gamma_{2,4}N_j + \gamma_{2,5}E_j + u_{2,j}$$

Our preferred theoretical model eliminates the effect of the latent rate on the random intercept, $\gamma_{0,1}$. We use the standard distributional assumptions for most multilevel SEM's, that is, the random effects are normally distributed with constant variance (across students, occasions, and teachers) and the standard assumptions for the covariance structure within (correlated) and across (uncorrelated) levels with two notable exceptions. First, the *e* and *r* within teacher residual random effects are assumed to be mutually uncorrelated. This follows from the fact that they pertain to conceptually different types of repeated measures within teachers, specific student test scores versus rates on a specific occasion. Second, the *u*'s are assumed to have a multivariate normal distribution in the base model but in our preferred theoretical model, u_0 vanishes, that is, the variance of u_0 goes to zero.

Authors' Note

The opinions expressed are those of the authors and do not represent the views of the Institute of Education Sciences or the U.S. Department of Education.

Declaration of Conflicting Interests

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: Drs. Ben Clarke, Scott Baker, and Hank Fien are eligible to receive a portion of royalties from the University of Oregon's distribution and licensing of certain ELM-based works. Potential conflicts of interest are managed through the University of Oregon's Research Compliance Services. An independent external evaluator and coauthor of this publication completed the research analysis described in the article.

Funding

The authors 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.

ORCID iD

Keith Smolkowski https://orcid.org/0000-0003-2565-3297

References

Agodini, R., & Harris, B. (2010). An experimental evaluation of four elementary school math curricula. *Journal of Research on Educational Effectiveness*, *3*, 199–253. doi:10.1080/19345741003770693

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

Asparouhov, T., & Muthén, B. (2012). *Mplus. Technical appendices: Saddle points* (Version 7). Retrieved from https://www.statmodel.com/download/SaddlePoints2.pdf

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*, *44*, 48–57. doi:10.1016/j.ecresq.2011.09.004

Barnes, M. A., Klein, A., Swank, P., Starkey, P., McCandliss, B., Flynn, K., . . . Roberts, G. (2016). Effects of tutorial interventions in mathematics and attention for low-performing preschool children. *Journal of Research on Educational Effectiveness*, *1*, 155–178. doi:10.1080/19345747.2016.1191575

Clarke, B., Baker, S., Smolkowski, K., Doabler, C. T., Strand Cary, M., & 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

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

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

Clements, D. H., Agodini, R., & Harris, B. (2013). *Instructional practices and student math achievement: Correlations from a study of math curricula* (NCEE Evaluation Brief No. 2013-4020). Washington, DC: National Center for Educational Evaluation and Regional Assistance, Institute of Education Sciences.

Connor, C. M., Morrison, F. J., Fishman, B. J., Ponitz, C. C., Glasney, S., Underwood, P. S., . . . Schatschneider, C. (2009). The ISI classroom observation system: Examining the literacy instruction provided to individual students. *Educational Researcher*, *38*, 85–99. doi:10.3102/0013189X09332373

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

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.

Doabler, C.T., Nelson, N. J., Kennedy, P. C., Stoolmiller, M., Fien, H., Clarke, B., . . . Baker, S. K. (in press). Investigating the longitudinal effects of a core kindergarten mathematics program on teachers' use of evidence-based teaching practices. *Learning Disability Quarterly*.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*, 1428–1446. doi:10.1037/0012-1649.43.6.1428

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

Ginsburg, H. P., & Baroody, A. J. (2003). *Test of Early Mathematics Ability—Third Edition (TEMA-3)*. Austin, TX: Pro-Ed.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*, 81–112. doi:10.3102/003465430298487

Heyman, R. E., Lorber, M. F., Eddy, J. M., & West, T. V. (2014). Behavioral observation and coding. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd ed., pp. 345–372). New York, NY: Cambridge University Press.

Ho, A. D., & Kane, T. J. (2013). *The reliability of classroom observations by school personnel* (MET Project Research Paper) Seattle, WA: Bill & Melinda Gates Foundation.

Hughes, C. A., Morris, J. R., Therrien, W. J., & Benson, S. K. (2017). Explicit instruction: Historical and contemporary contexts. *Learning Disabilities Research & Practice*, *32*, 140–148. doi:10.1111/ldrp.12142

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. doi:10.1037/a0014939

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.

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:10.3102/0162373714536608

Muthén, L. K., & Muthén, B. O. (2015). *Mplus user's guide* (Version 7.31). Los Angeles, CA: Author.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher*, *38*, 109–119. doi:10.3102/0013189X09332374

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2016). Multilevel structural equation models for assessing moderation within

and across levels of analysis. *Psychological Methods*, *21*, 189–205. doi:10.1037/met0000052

Raudenbush, S. W. (2008). Advancing educational policy by advancing research on instruction. *American Educational Research Journal*, *45*, 206–230. doi:10.3102/0002831207312905

Rothman, A. (2013). Exploring connections between moderators and mediators: Commentary on subgroup analyses in intervention research. *Prevention Science*, *14*, 189–192. doi:10.1007/s11121-012-0333-y

Sood, S., & Jitendra, A. K. (2013). An exploratory study of a number sense program to develop kindergarten students' number proficiency. *Journal of Learning Disabilities*, *46*, 328–346. doi:10.1177/0022219411422380

Wang, A. H., Firmender, J. M., Power, J. R., & Byrnes, J. P. (2016). Understanding the program effectiveness for prekindergarten and kindergarten environments: A meta-analytic review. *Early Education and Development*, *27*, 692–713. doi:10.1080/10409289.2016.1116343