Journal of Educational Psychology

Differential Effects of the Classroom on African American and Non–African American's Mathematics Achievement

Katerina Schenke, Tutrang Nguyen, Tyler W. Watts, Julie Sarama, and Douglas H. Clements

Online First Publication, March 23, 2017. http://dx.doi.org/10.1037/edu0000165

0022-0663/17/$12.00 http://dx.doi.org/10.1037/edu0000165

Differential Effects of the Classroom on African American and Non–African American's Mathematics Achievement

Katerina Schenke
University of California, Los Angeles

Tutrang Nguyen and Tyler W. Watts
University of California, Irvine

Julie Sarama and Douglas H. Clements
University of Denver

We examined whether African American students differentially responded to dimensions of the observed classroom-learning environment compared with non–African American students. Further, we examined whether these dimensions of the classroom mediated treatment effects of a preschool mathematics intervention targeted at students from low-income families. Three observed dimensions of the classroom (teacher expectations and developmental appropriateness; teacher confidence and enthusiasm; and support for mathematical discourse) were evaluated in a sample of 1,238 preschool students in 101 classrooms. Using multigroup multilevel mediation where African American students were compared with non–African American students, we found that teachers in the intervention condition had higher ratings on the observed dimensions of the classroom compared with teachers in the control condition. Further, ratings on teacher expectations and developmental appropriateness had larger associations with the achievement of African American students than for non–African Americans. Findings suggest that students within the same classroom may react differently to that learning environment and that classroom learning environments could be structured in ways that are beneficial for students who need the most support.

Educational Impact And Implications Statement

Data from this study come from a randomized control trial of a preschool mathematics intervention. We used multigroup multilevel structural equation modeling to test whether certain instructional practices explained why African American students had differential gains in mathematics achievement as compared with non-African American students. This study suggests that, on average, African American students benefit differently from certain instructional practices (e.g., teacher expectations and developmental appropriateness) more than non-African American students. Implications suggest that content-specific interventions could be designed to include changing teachers' beliefs about students' abilities.

Keywords: achievement gap, classroom observations, mathematics learning, multilevel structural equation modeling

Supplemental materials: http://dx.doi.org/10.1037/edu0000165.supp

Identifying ways to promote student learning, especially for marginalized groups, is an important goal of educational research. On average, the mathematics achievement gap between low-income minority students and high-income White students is close to two thirds of a standard deviation at the start of kindergarten (Duncan & Magnuson, 2005; Loeb & Bassok, 2008; Reardon & Robinson, 2008). This has led both researchers and educational advocacy groups to call for high-quality mathematics instruction prior to school-entry, as such efforts could reduce achievement gaps at the beginning, and throughout, K–12 schooling (National

Katerina Schenke, Graduate School of Education & Information Studies, University of California, Los Angeles; Tutrang Nguyen and Tyler W. Watts, School of Education, University of California, Irvine; Julie Sarama and Douglas H. Clements, Morgridge College of Education, University of Denver.

This research was supported by the Institute of Education Sciences, U.S. Department of Education through Grants R305A120813, R305K05157, and R305A110188 and the NICHD-supported Irvine Network on Interventions in Development (HD065704 P01). The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education. The authors express appreciation to the school districts, teachers, and children who participated in this research. We also thank Greg Duncan and Sandra Graham for comments on earlier drafts of this article and Erik Ruzek and Kevin Schaaf for their consultation on the analysis.

Correspondence concerning this article should be addressed to Katerina Schenke, Graduate School of Education & Information Studies, University of California, Los Angeles, CA 90095-1521. E-mail: kschenke@ucla.edu

Council of Teachers of Mathematics (NCTM), 2000; National Mathematics Advisory Panel, 2008; National Research Council, 2009).

The most critical feature of a high-quality educational environment is a knowledgeable and responsive adult (Darling-Hammond, 1997; Ferguson, 1991; National Research Council, 2009; Sarama, & DiBiase, 2004; Schoen, Cebulla, Finn, & Fi, 2003). This may be especially true for classroom-based interventions targeted at closing racial achievement gaps, as certain instructional practices may be more important for some racial/ethnic groups than others (e.g., Bodovski & Farkas, 2007; Sonnenschein & Galindo, 2015; Wenglinsky, 2004). However, understanding which specific elements of the classroom, and whether students within the same classroom respond differently to the same classroom practices, has been largely unexplored. This is an important endeavor as promoting the achievement for marginalized groups, notably African American children, remains scarce (Stinson, 2006).

Data for this study come from a randomized control trial of an intervention that was shown to have a larger effect—almost double in magnitude—on the average mathematics achievement of African American students than of non–African American students (Clements, Sarama, Spitler, Lange, & Wolfe, 2011). Analyses suggest that classroom quality—measured broadly—mediated this effect (Clements et al., 2011). However, it is unknown what specific processes within the classroom led to this larger treatment effect. The goal of the current paper is to explore whether specific classroom practices could explain why African American students, on average, fared better than non–African American students at the end of the intervention. The results of our analyses have broader implications for designing classroom-learning environments to support the learning of marginalized students.

Inequality in Early Learning Experiences

Children with low early academic achievement and those who encounter early learning problems face continuing negative consequences that accumulate over time (Alexander, Entwisle, & Horsey, 1997; Brooks-Gunn & Duncan, 1997; Duncan, Brooks-Gunn, & Klebanov, 1994; Fryer & Levitt, 2006; Huston & Bentley, 2010). This is particularly salient for low-income, minority children who typically begin school with fewer academic skills than their middle- to high-income peers (Duncan & Magnuson, 2005; Lee & Burkam, 2002; Loeb & Bassok, 2008; Reardon & Robinson, 2008). Differences between the educational experiences of low-income and high-income students have been well documented in the literature (Aikens & Barbarin, 2008; Kozol, 1991; Oakes, 1990). However, racial/ethnic differences in students' access to high quality learning environments still exist even when socioeconomic factors are controlled (Lubienski, 2002; Oakes, 1990). In general, race and socioeconomic status are highly correlated, making the study of the unique contributions of race and socioeconomic status difficult to untangle.

Even when minority students have access to the same schools and classroom environments, inequalities exist in how students of different minority groups experience the classroom environment. For example, teachers have been shown to have differential expectations for students of different racial/ethnic groups even when the previous achievement of these students was equivalent (McKown & Weinstein, 2008). Further, when looking within the same functional learning environment, the effects that certain instructional practices have on students have been found to differ according to racial/ethnic group. For example, African Americans on average, compared with students of other races/ethnicities, have been found to differentially benefit from certain instructional practices such as collaborative problem solving (Lubienski, 2006), an emphasis on specific mathematics content (Bodovski & Farkas, 2007; Sonnenschein & Galindo, 2015; Wenglinsky, 2004), and more opportunities to learn higher-level mathematics such as reasoning and problem solving (Battey, 2013; Bodovski & Farkas, 2007; Boaler, 1998; Ladson-Billings, 1997; Lubienski, 2002; Silver & Stein, 1996).

Unfortunately, we still lack a thorough understanding of why African American students may differentially benefit from certain instructional practices compared with students from other racial/ethnic groups. Some scholars suggest that African American students are oriented toward different learning styles (Berry, 2003; Ladson-Billings, 1997) and that these different styles of learning should be considered during teaching. García Coll and colleagues (1996) proposed an integrative model for studying child development and suggested that elements such as the learning environment and culture influence students' psychological experiences. Even when children are in the same environment, their interpretations of certain aspects of the environment may vary according the child's previous cultural experiences, values, and goals. Differences in students' classroom perceptions have been observed in empirical work that has shown that African American students' attitudes toward learning may differ from those of White students (Lubienski, 2002; Strutchens & Silver, 2000). This broad theory proposed by García Coll and colleagues (1996) outlines possible mechanisms for understanding the experiences of minority students, but no theory to date specifically addresses students' differential responsiveness to classroom practices.

Dimensions of the Classroom Learning Environment

Early childhood environments have been described along a variety of dimensions (see Stipek & Byler, 2004 for a review), such as teacher–student interactions (Abbott-Shim, Lambert, & McCarty, 2000; Harms, Clifford, & Cryer, 1998; Pianta & Hamre, 2009; Stipek & Byler, 2004; NICHD ECCRN, 1996), quality of instruction (Abbott-Shim, Lambert, & McCarty, 2000; Pianta & Hamre, 2009), types of activities done in the classroom (Harms et al., 1998), behavior management (Abbott-Shim, Lambert, & McCarty, 2000; Pianta & Hamre, 2009; Stipek & Byler, 2004), and others. Empirically, observational measures are typically used to examine associations between these dimensions and children's developmental outcomes (e.g., Bryant, Burchinal, Lau, & Sparling, 1994; Peisner-Feinberg & Burchinal, 1997). In the present study, we described the classroom environment as (a) teachers' expectations and the developmental appropriateness of their instruction (developmental appropriateness as operationalized by the expectations teachers have of what preschool students are capable of learning), (b) teachers' confidence and enthusiasm in their teaching, and (c) teachers' support for mathematical discourse.

Teacher Expectations

The expectations teachers have of their students may have an effect on students' learning (e.g., Brophy, 1986; Gill & Reynolds, 1999; Rosenthal & Jacobson, 1968) and are related to high quality and equitable instruction (Askew, Brown, Rhodes, Wiliam, & Johnson, 1997; Clarke, Frazer, DiMartino, Fisher, & Smith, 2003; Clements & Sarama, 2007, 2008; NCTM, 2000). These expectations influence whether teachers decide to provide or constrain the opportunities they give their students (Brophy & Good, 1970). For example, a teacher who has high expectations for a student may give that student more opportunities to answer questions during class or further press the student to explain their thinking. In contrast, a teacher who has low expectations for a student may not give that student enough time to respond to a question or may not even call on that student in the first place, thereby making the student miss a key learning opportunity. Many teachers hold low expectations for students whom they believe have low ability or achievement (Brophy & Good, 1970), and these low expectations are often related to the ethnicity of the student (Dusek & Joseph, 1983; Jussim, Eccles, & Madon, 1996; Madon, Jussim, Keiper, Eccles, Smith, & Palumbo, 1998; McKown & Weinstein, 2008). Indeed, teachers who hold particularly low academic expectations for African American students spend more time on behavioral corrections than content instruction (Gill & Reynolds, 1999; Jussim, 1989; Jussim et al., 1996; Kuklinski & Weinstein, 2001; Madon et al., 1998; Raudenbush, 1984; Steele, 1997).

Teacher expectations are often measured at the student level, whereby teachers are asked to complete surveys regarding specific students in their classroom (e.g., McKown & Weinstein, 2008). As such, this method does not readily allow investigation of between-teacher levels of expectations where comparisons across teachers' expectations for their students can be made. One exception is Rubie-Davies (2007), who created categories (low, average, and high) of teachers from their individual ratings of students such that teachers' expectations of students are nested within the teacher. However, this method may mask variation between teachers' expectations. If one imagines teachers' ratings of expectations for each student spanning a normal distribution, a teacher's ratings will vary about the mean. However, once these ratings are aggregated to the teacher-level, thereby calculating a teacher-level rating of expectations, differences *between* teacher's ratings will appear small. Furthermore, understanding differences between teachers' expectations of their students leads to questions about whether interventions can be designed to change these mean-level expectations. This topic has not been exhaustively explored through the implementation of interventions (for exceptions see Proctor, 1984; Rubie-Davies, Peterson, Sibley, & Rosenthal, 2015; Weinstein et al., 1991).

As an exception, Weinstein and colleagues (1991) designed an extensive intervention to change the school and classroom climate to improve low achieving ninth grade students' academic performance and behavior. Whereas one component of this intervention was to raise teachers' expectations of their students, other elements of the classroom-learning environment were also targeted, such as adapting readings from the honors English courses and using heterogeneous grouping strategies, thereby holding low achieving students to higher standards. In another example, Rubie-Davies and colleagues (2015) designed an intervention providing teachers with professional development (PD) aimed at increasing the expectations teachers had of their students. Even though the intervention was described as one focusing on teacher expectations, the PD also focused on other elements of the classroom climate such as increasing student motivation, providing useful feedback, and providing opportunities for promoting student autonomy. Therefore, the effects of the intervention on students' academic achievement could be attributed to all of these classroom dimensions. Whereas previous studies (e.g., Rubie-Davies et al., 2015; Weinstein et al., 1991) hypothesized that part of their observed academic gains were attributable to changes in teachers' classroom behaviors, the present study is the first to empirically test these hypotheses by directly measuring elements of the teacher's classroom practices and by conducting mediational analyses.

Understanding what scholars and practitioners consider as developmentally appropriate in the early years has been questioned. For example, some argue that intensive content-specific instruction in preschool can provide the basic foundational skills that can help prepare children for the academic nature of kindergarten (Ginsburg, Inoue, & Seo, 1999; Seo & Ginsburg, 2004). Others argue that subjecting children to harsh forms of instruction and imposing material they are not ready to learn can be detrimental, forcing young children to engage in developmentally inappropriate forms of drills and practices in mathematics (Bishop-Josef & Zigler, 2011). Young children can indeed engage in various types of mathematical activity if provided with the appropriate opportunities to do so; however, teachers often fail to provide such opportunities to their students or may believe that such practices are developmentally inappropriate (Dunn & Kontos, 1997; Hitz & Wright, 1988). A central premise of the current study is that the developmental needs of students are met when students are held to high expectations, that is, when teachers believe students can engage in higher-level thinking. Preschool children—especially low-income, minority children—have the potential to learn and acquire math skills and concepts (Clements, Baroody, & Sarama, 2013; Ginsburg, Lee, & Boyd, 2008). If teachers and those who work with teachers underestimate what children already know and can learn, they will not present appropriate and challenging mathematics activities.

Teachers' Confidence and Enthusiasm

The beliefs teachers have about their teaching influence their instructional practices (Pajares, 1992; Stipek, 1998; Thompson, 1984) and, in turn, their students' achievement (Evertson, Emmer, & Brophy, 1980). Descriptions of effective mathematics classrooms likely include a teacher who sparks joy and excitement in her students. The ability for a teacher to instruct in a confident and enthusiastic manner is thought to instill confidence and enthusiasm in students and therefore promote learning and achievement. Though confidence and enthusiasm are thought to be important dimensions of the classroom climate, these attributes are rarely measured in quantitative studies of instructional practices.

One exception is an early study by Evertson and colleagues (1980) who conducted observations of less- and more-effective teachers as measured by student achievement on a state standardized test, and found differences in observed enthusiasm for teaching and confidence between the two groups of teachers. Another exception is a study by Stipek and colleagues (2001), who measured teacher confidence and enthusiasm through observations and teacher self-report. They found that teacher's self-reported confidence in teaching mathematics was correlated with the students' classroom average self-confidence in mathematics. Further, they also reported that teachers' enjoyment of mathematics correlated with students' enjoyment of mathematics. Although theory would support the association between teachers' confidence and enthusiasm and student achievement (Stipek, Givvin, Salmon, & MacGyvers, 2001), the authors are not aware of any studies directly testing this association. Evertson and colleagues (1980) only provide weak evidence of such an association as these practices were not directly correlated with student achievement, but rather were only investigated in the context measuring a subsample of teachers chosen by the researchers to examine differences between more effective and less effective teachers.

Support for Mathematical Discourse

Support for mathematical discourse has been recognized as an important practice for fostering students' mathematical learning (Hiebert, & Grouws, 2007; Huttenlocher, Vasilyeva, Waterfall, Vevea, & Hedges, 2007; NCTM, 1991). For example, allowing students the opportunity to explain their thinking, elaborate on concepts, and generate mathematical talk, have been identified as high-level mathematical activities (Henningsen & Stein, 1997; NCTM, 1991) and these practices have effects on students' mathematical knowledge development (Walshaw & Anthony, 2008). Such practices are thought to help students increase their understanding of mathematics by helping students better internalize mathematical content and by allowing other students to learn from hearing student-generated explanations (Chi, 2000). Additionally, when students generate their own mathematical explanations, teachers are able to tailor their instruction to address inaccuracies in explanations or misconceptions (Franke, Fennema, & Carpenter, 1997). However, many teachers do not talk to their students about mathematics, even when the student initiates the mathematical talk. In one study, when students made a mathematical utterance, their teachers ignored it 60% of the time and only responded mathematically 10% of the time (Diaz, 2008).

Some scholars promote a conceptual and problem-solving approach infrequently emphasized in schools serving low-income children (Stipek & Ryan, 1997) that may explicitly support African American students' participation in increasingly sophisticated forms of mathematical communication and argumentation. For example, asking students "How do you know?" as opposed to a more didactic approach of giving information [frequently used with African American students (Haberman, 1991; Jackson & Wilson, 2012; Ladson-Billings, 1997)] may be especially beneficial for African American students.

The Present Study

Data for the current analysis were drawn from an evaluation of the "Technology-enhanced, Research-based, Instruction, Assessment, and professional Development" (TRIAD) model for scaling up successful interventions (Clements et al., 2011; Sarama & Clements, 2013; Sarama et al., 2008). This study assessed an instantiation of the TRIAD model that implemented the *Building Blocks* curriculum (Clements & Sarama, 2008), an empirically validated preschool mathematics program based on developmentally appropriate learning trajectories (see Clements & Sarama, 2008). These learning trajectories were operationalized by a series of "developmental progressions," which identified a given topic or domain. To attain a certain mathematical competence in a given topic or domain, students are taught each successive level in a developmental progression using research-based tasks and instructional strategies.

Building Blocks emphasizes both numeracy and spatial/geometric concepts and procedures through the use of whole and small group activities. In addition to the curriculum, the TRIAD intervention included a software package and extensive teacher support, in the form of training on the software, 13 professional development (PD) sessions before and during the school year, and study-appointed mentors and coaches. During the PD sessions, the training addressed numerous effective teaching practices, including how to increase the cognitive demand of the mathematics taught, how to use formative assessment, and how to encourage cognitive development through the use of learning trajectories (see Clements et al., 2011). Of particular importance to this study, PD administrators also worked with treatment teachers to change their beliefs about the ability of low-income and minority children to learn advanced mathematics.

The goal of the present study was to understand (a) which classroom dimensions were associated with African American and non–African American students' mathematics achievement, and (b) whether classroom dimensions mediated the association between the intervention and students' mathematics achievement at the end of preschool. We focused on the dichotomy of African American and non–African American students because treatment evaluations of the intervention suggest that although Hispanic students made statistically significant gains in mathematics achievement, these gains were not statistically significantly different than those of White students (Clements et al., 2011).

A critical feature of the present study is that we were able to examine the associations of specific dimensions of the classroom (e.g., teacher expectations and developmental appropriateness, teacher confidence and enthusiasm, and high quality instructional practices) within the context of a randomized control trial. Further, we examined whether these dimensions of classroom practices were associated with later measures of student achievement.

Previous research affords the following hypotheses and predictions: We expect teacher expectations and developmental appropriateness will matter more for African American students' achievement than for non–African American students' achievement and help explain (i.e., mediate) the association between the TRIAD treatment and African American students' mathematics achievement (H1). Teacher confidence and enthusiasm will be associated with students' posttest mathematics achievement—for both African American and non–African American students—and will be a mediator of the treatment intervention on mathematics achievement (H2). However, in the absence of theory, we do not predict any differences in this association between African American and non–African American students. Finally, we predict that support for mathematical discourse will be statistically significantly associated with mathematics achievement for both groups (African American *and* non–African American students) but, in the absence of theory, will not differentially predict between the groups (H3). These hypotheses were tested using a multigroup multilevel mediation model conducted for each of the three mediators (teacher expectations and developmental appropriateness, teacher confidence and enthusiasm, and high quality instructional practices).

Method

Participants and Procedure

Data for the current analysis were drawn from the TRIAD evaluation—a study that assessed the scale-up and student-level impacts of the TRIAD intervention model, of which a key component was the *Building Blocks* mathematics curriculum (see Clements et al., 2011; Clements, Sarama, Wolfe, & Spitler, 2013; Sarama, Clements, Wolfe, & Spitler, 2012). The TRIAD evaluation recruited 42 low-resource schools in two states (New York and Massachusetts) to participate in the scale-up evaluation of *Building Blocks*. Schools were grouped into eight blocks based on state achievement scores, and then randomly assigned within block to one of two conditions: (a) *Building Blocks* curriculum condition (treatment), or (b) business-as-usual (control).[1] In the following school year (2006–2007), student-level data collection began, and 1,375 preschool students attending the 42 study-schools were recruited for study participation at the beginning of the school year.

The majority of students in the study qualified for free or reduced price lunch (84% of those for whom we had available data for), and 55% identified as African American and 22% as Hispanic. The analysis sample only included students who had valid baseline and posttreatment test score data (n = 1,305 students), and were in classrooms that were observed on the observation protocol resulting in a final sample size of 1,238 students in 101 classrooms (approximately 70% of the students in each classroom). Twenty-nine percent of the sample was in the control group (n = 350 students), and 71% of the sample was in the treatment group (n = 888 students). Table 1 shows the descriptive statistics for both student- and classroom-level variables for the overall sample and by group (African American and non–African American students).

Student-Level Measures

Mathematics achievement. Children's mathematical knowledge was assessed at preschool entry (pretest) and at the end of the preschool year (posttest) using the Research-based Elementary Math Assessment (REMA; Clements, Sarama, & Liu, 2008). The REMA was designed to assess the mathematics knowledge of children from ages three to eight, specifically students' number (e.g., object counting, number comparison, numeral recognition) and geometry (e.g., shape identification, measurement, patterning) skills, and was administered in two one-on-one sessions where a research assistant verbally asked participants to respond to each item. After the child incorrectly answered four items in a row, the assessment stopped. The assessments were videotaped and later coded by a team of trained researchers for correctness and strategy use. Using these dichotomous scores, Rasch analysis—where item difficulty is considered in estimating a child's overall score—was employed. Using data from multiple years, the scores were then placed on a vertical scale using data from students in grades preschool, kindergarten, first grade, and second grade where first grade was used as the benchmark. Final scores were scaled to have a mean of 0 and a standard deviation of 1.

The REMA has been validated across three diverse samples of preschool-aged children and produced an overall reliability of .93 (see Clements et al., 2008). Further, it has been shown to have a .86 correlation with another empirically validated measure of early mathematics achievement, the Child Math Assessment: Preschool Battery (Klein, Starkey, & Wakeley, 2000) and a .74 correlation with the Woodcock-Johnson Applied Problems subtest (Clements et al., 2008). In the current sample, the measure had a reliability (alpha) of .92.

Student-level covariates. Information regarding student ethnicity, gender, age, and limited English proficiency status were collected from the schools at the beginning of the preschool year.

Classroom-Level Measures

Classroom observations. Live classroom observations were conducted twice during the year using the Classroom Observation of Early Mathematics Environment and Teaching (COEMET; Clements & Sarama, 2000/2016). Observation ratings from both time points (the first observation was conducted between October and December and the second observation between March and May of the school year) were combined to obtain an average rating of the classroom environment during the school year. Observers spent about half a day in each classroom from before the children arrived until right before lunch and were blind to experimental condition. The COEMET is divided into two sections. The first section asked observers to document the number of computers in the classroom and the start and end time of all activities that took place during the observation using interval coding. The second section of the COEMET included broader measures rated during each mathematics activity and contained 28 indicators measuring constructs such as *supporting children's conceptual understanding, teaching strategies,* and *expectations*. Observers coded each item on a 5-point Likert scale ranging from *strongly disagree* (1) to *strongly agree* (5). Ratings for each mathematics activity were averaged to obtain a mean score across the 28 indicators. Interrater reliability for the COEMET, computed via simultaneous classroom visits by pairs of observers (10% of all observations, with pair memberships rotated), was 88% (i.e., 88% of the 28 Likert items were coded the same by both assessors); of the 12% of disagreements, 99% were of the same polarity (i.e., if one was agree, the other was strongly agree). Coefficient alpha (interitem correlations) for the two in-

[1] The study actually employed three conditions, two treatment conditions and one control condition. However, during the preschool year, the two treatment conditions did not differ, and thus were combined into one condition in the current analysis.

Table 1
Descriptives for Student-Level and Classroom-Level Analysis Sample (N = 1,238 Students in 101 Classrooms)

Level of analysis	Overall sample (*N* = 1,238)				African American (*N* = 675)				Non–African American (*N* = 563)			
	Mean	*SD*	Min	Max	Mean	*SD*	Min	Max	Mean	*SD*	Min	Max
Student-level												
Pre-math	−3.23	.82	−7.2	−.4	−3.31	.77	−7.2	−.6	−3.13	.88	−7.2	−.4
Post-math	−1.97	.70	−4.9	.6	−2.10	.68	−4.9	.6	−1.82	.70	−4.8	.2
Age	4.34	.35	3.7	6.7	4.27	.33	3.7	5.8	4.41	.36	3.7	6.7
African American	55%				—				—			
White	18%				—				39%			
Hispanic	22%				—				48%			
Other ethnicity	5%				—				13%			
LEP	17%				3%				33%			
Special education	17%				15%				20%			
Male	49%				50%				49%			

	Overall sample (*N* = 101)				Treatment (*N* = 888 students in 69 classrooms)				Control (*N* = 350 students in 32 classrooms)			
	Mean	*SD*	Min	Max	Mean	*SD*	Min	Max	Mean	*SD*	Min	Max
Classroom-level												
No. computers	2.06	1.32	0.0	6.0	2.45	1.21	0.0	6.0	1.22	1.16	0.0	4.0
No. math activities	6.97	2.85	1.0	15.0	7.75	2.64	2.0	15.0	5.28	2.58	1.0	12.0
Minutes of math	30.90	14.60	8.2	92.5	32.39	15.39	8.2	92.5	27.68	12.33	10.3	60.7
Homework	2.40	1.06	1.0	4.0	2.51	1.02	1.0	4.0	2.19	1.12	1.0	4.0
Classroom size	17.07	3.18	10.0	25.0	17.48	2.98	11.0	25.0	16.22	3.45	10.0	22.0
Master's degree	0.82	0.39	0.0	1.0	0.88	.33	0.0	1.0	0.69	0.47	0.0	1.0
No. years teaching	15.29	8.61	0.0	35.0	15.10	8.35	1.0	32.0	15.69	9.27	0.0	35.0
PD hours	14.90	8.49	0.0	342.0	15.10	8.35	0.0	32.0	14.45	8.93	0.0	35.0
Percent African American	0.54	0.35	0.0	1.0	0.57	0.35	0.0	1.0	0.49	0.34	0.0	1.0
Percent SES	0.38	0.19	0.0	0.9	0.40	0.18	0.1	0.9	.33	0.19	0.0	0.7
Confidence	3.01	0.37	1.8	3.8	3.09	0.30	2.0	3.6	2.84	0.46	1.8	3.8
Expectations	3.15	0.21	1.9	3.8	3.19	0.16	2.4	3.8	3.09	0.27	1.9	3.4
Discourse	3.00	0.38	1.9	4.1	3.07	0.34	2.4	4.1	2.85	0.43	1.9	3.6

Note. Mathematics achievement was scored using Rasch analysis on a vertical scale. PD = professional development; SES = socioeconomic status; LEP = limited English proficient. PD hours is the number of hours of professional development before the intervention. No. computers is the number of computers. No. math activities is number of math activities. No. years teaching is number of years teaching. Difference in the number of computers and number of discrete mathematics activities were statistically significantly different across the treatment and control groups at $p < .001$.

struments ranged from .95 to .97 in previous research (Clements & Sarama, 2008; Clements et al., 2011).

Three dimensions of classroom quality were used in the analysis: *expectations and developmental appropriateness; teacher confidence and enthusiasm; and support for mathematical discourse.* Information on the factor analyses, items, and standardized factor loadings is available in the supplemental materials (Tables S1 through S3). Correlations among factors were moderate in magnitude: $r = .65$ between teacher confidence and enthusiasm and expectations and responsiveness to developmental needs; $r = .48$ between teacher confidence and enthusiasm and support for mathematical discourse; $r = .66$ between expectations and responsiveness to developmental needs and support for mathematical discourse.

Expectations and responsiveness to developmental needs. Eight items were used to describe the extent to which the teacher had high expectations for her students and whether or not the teacher engaged with children in a developmentally appropriate manner. Developmental appropriateness was operationalized by the researchers as the extent to which teachers engaged in practices that matched the developmental abilities and potential of the students. This meant that a teacher who was rated high on developmental appropriateness engaged students in higher-level thinking around mathematics and English language arts instead of engaging students in only play-based activities, with no emphasis on higher-level thinking. Standardized factor loadings ranged from .73 to .88 ($\alpha = .94$). We created a composite using items weighted by their standardized factor loadings.

Teacher confidence and enthusiasm. Three items were used to describe characteristics that were associated with how confident and enthusiastic the teacher appeared during class. Specifically, items described if the teacher was confident in her teaching, showed interest and value in her students, and displayed enthusiasm for mathematics. Standardized factor loadings ranged from .68 to .84 ($\alpha = .82$). We created a composite score using items weighted by their standardized factor loadings for the analysis.

Support for mathematical discourse. Eight items characterized the specific mathematics practices and strategies the teacher used during the observed lessons. These items captured the extent to which the teacher elicited higher-order thinking in mathematics through supporting students' explanations and thinking. Standardized factor loadings ranged from .82 to .88 ($\alpha = .94$). We created a composite using items weighted by their standardized factor loadings.

Classroom-level covariates. Classroom-level covariates came from three sources: (a) the teacher survey, (b) classroom observations, and (c) student demographic information. Study teachers were administered a survey twice during the school year, in the fall and spring of the study year. For teachers who did not respond to the fall survey (two teachers), we used their responses from the spring survey. We included information on whether the teacher held a master's degree, the number of students in the classroom, the number of years of preschool through Grade 12 teaching experience, the number of PD hours prior to the intervention, and whether the teacher assigned mathematics homework measured on a one to four Likert scale with anchor points *almost never* to *a lot*. From the classroom observations, we controlled for the number of computers in the classroom, number of minutes spent on mathematics, and the number of discrete mathematics tasks observed. Number of minutes spent on mathematics was the observed actual time of mathematics instruction from the recorded start and end times on the COEMET. The number of discrete mathematics tasks observed was obtained from the interval coding of the COEMET where observers had to start coding at the beginning of each mathematics activity.

Finally, we aggregated information from the student demographic information to include percent African American students in the classroom as well as percent of students' whose mother's education was college or higher as a proxy for classroom-level socioeconomic status (SES). Because of constraints on researcher resources, only data from approximately 70% of the students in each classroom were collected. As such, the classroom-aggregated measures serve as approximations of the average percent of African American students and average SES of the classroom. The racial/ethnic breakdown of the classroom composition varied considerably. The mean percent of African American students in a classroom was 54% ranging from classrooms with no African American students (0%) to classrooms with only African American students (100%).

Overview of Analytic Models and Method

To examine differences in mediational paths between African American and non–African American students, we conducted a multigroup multilevel meditational analysis (Asparouhov & Muthén, 2012; Retelsdorf, Schwartz, & Asbrock, 2015). With this method, we combined a multilevel structural equation model examining mediation effects (Preacher, Zyphur, & Zhang, 2010) with multigroup analysis in *Mplus 7.2* (Muthén & Muthén, 2013). We specified a 2→2→1 model of mediation in which classroom observation (level 2) mediated the association between treatment (level 2) and student mathematics achievement (level 1). The equations for the direct effect (path c) of the treatment on post mathematics achievement are:

Level-1 (student-level) equation:

$$PostMath_{ij} = \beta_{0j} + \beta_1 PreMath_{ij} + \lambda_2 Covariates_{ij} + r_{ij}$$

Level-2 (classroom-level) equation:

$$\beta_{0j} = \gamma_{00} + \gamma_c Treatment_j + \lambda_2 Covariates_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

where ($PostMath_{ij}$) is the posttreatment measure of mathematics achievement for the ith student in school j. Posttreatment mathematics achievement is modeled as a function of an overall intercept (γ_{00}, the main effect for the treatment group ($\gamma_c Treatment_j$), a vector of level-2 covariates ($\gamma_2 Covariates_j$), the students' baseline mathematics achievement ($\beta_1 PreMath_{ij}$), a vector of level one covariates ($\gamma_2 Covariates_j$), and a classroom-level (u_{0j}) and student-level (r_{ij}) error term.

The equations for the effect of the mediator on posttreatment mathematics achievement (path b) are the same as above, except that a mediator component is added to the level-2 equation:

Level-1 (student-level) equation:

$$PostMath_{ij} = \beta_{0j} + \beta_1 PreMath_{ij} + \lambda_{2j} Covariates_{ij} + r_{ij}$$

Level-2 (classroom-level) equation:

$$\beta_{0j} = \gamma_{00} + \gamma_{c'} Treatment_j + \gamma_b Mediator_j + \lambda_3 Covariates_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

The equation for the effect of the treatment on the mediator (path a) is:

$$Mediator_j = \beta_0 + \beta_a Treatment_j + r_j$$

where the mediator is a function of an overall intercept (β_0), the main effect of the treatment ($\beta_a Treatment_j$), and a classroom-level error term (r_j).

Mediation is implied if the treatment is significantly related to the mediator (classroom observation), the mediator is significantly associated with the outcome (student achievement), and the mediator accounts for a significant portion of the variance of the association of the treatment on the outcome. Further, we also specified a multiple group model whereby the effect of classroom observation (level 2) on student mathematics achievement (level 1) differed between African American and non–African American students. Because our grouping variable varied within classroom, calculating varying effects for African American and non–African American students cannot be directly specified because the estimated variance/covariance matrix for the observed variables will be group-specific (i.e., classroom specific). However, the effect of the classroom observation on mathematics achievement is correlated between racial/ethnic groups because these students are in the same classroom, but the effect need not be the same for both ethnicities.

To circumvent this within the multilevel structural equation framework, we used a method proposed by Asparouhov and Muthén (2012) where latent variables are introduced in the model to account for the covariance between the group specific classroom effects (please refer to the supplemental materials for our *Mplus* code). Specifically, known latent classes using race (African American and non–African American) were specified in the model such that separate effects for African American and non–African American students are estimated. In the between portion of the model where level 2 effects are specified, two latent variables (our known latent classes) are introduced to represent the between level random effects for $e_{non\text{–}African\ American}$ and $e_{African\ American}$. The math outcome was specified to have a zero residual variance structure as well as loadings equal to one in both groups. The random effects were correlated within classroom to account for nesting. This results in a model where the outcome (posttreatment mathematics achievement) was represented by $e_{non\text{–}African\ American}$

and $e_{African\ American}$. A Wald test was used to examine differences in the *b* paths (path between classroom dimension and math achievement) between the two groups. To estimate the indirect effects and report a statistical test for mediation, we used a parametric bootstrap method (Efron & Tibshirani, 1986) in which the parameter point estimates for the indirect effect are generated from random draws of the parameter distributions for the *a* and *b* paths of the mediation model. This method has advantages over the Sobel test in that the distribution of the indirect effect is not assumed to be normally distributed (Preacher et al., 2010). To estimate the parametric bootstrap, we used a Web-based tool to generate R code developed by Selig and Preacher (2008) specifying a confidence interval of 95% and 20,000 random draws. The *p* values for the average indirect effects were then calculated directly in R and reported. In total, three models were run—one for each mediator.

Results

We first present results from the correlations among the variables of interest, which are displayed in Table 2. We then discuss results from each of the mediation models with the three classroom observation variables for non–African American and African American groups. These are displayed as figures (Figures 1 through 3) depicting path diagrams as well as in table form (Tables 3 through 5). We have also included information on the same mediation models comparing White students with Hispanic students ($n = 488$) to further justify why we collapsed these two racial/ethnic groups into the non–African American group. These results are described at the end of each of the mediator sections, and full tables can be found in the supplemental materials (Tables S4 through S6). Wald tests were only performed if the path from the classroom dimension to students' mathematics achievement was statistically significant for at least one of the groups.

Baseline mathematics achievement and posttreatment mathematics achievement were highly correlated, $r = .57$, $p < .001$. Across the whole sample, teacher confidence and enthusiasm was statistically significantly correlated with post mathematics achievement ($r = .18$, $p < .001$), as was expectations and responsiveness to developmental needs ($r = .09$, $p < .01$), but, support for mathematical discourse and posttest mathematics achievement were not significantly correlated, $r = -.03$, $p = .29$.

Number of computers in the classroom (recall that the *Building Blocks* intervention included software) was statistically significantly correlated with teacher confidence and enthusiasm and expectations and responsiveness to developmental needs ($r = .22$, $p < .001$; $r = .14$, $p < .001$, respectively) but not with support for mathematical discourse, $r = -.02$, $p = .48$. Number of mathematics activities was also significantly correlated with teacher confidence and enthusiasm and expectations and responsiveness to developmental needs ($r = .31$, $p < .001$; $r = .11$, $p < .001$, respectively) but significantly negatively correlated with support for mathematical discourse, $r = -.26$, $p < .001$. Whether the teacher assigned mathematics homework was significantly negatively correlated with all three classroom-observation variables ($r = -.08$, $p < .01$; $r = -.10$, $p < .001$; $r = -.28$, $p < .001$; for teacher confidence and enthusiasm, expectations and responsiveness to developmental needs, and support for mathematical discourse, respectively).

Mediator 1: Expectations and Responsiveness to Developmental Needs

Figure 1 displays the path diagram for the first mediator: expectations and responsiveness to developmental needs. The treatment had a positive impact on teacher expectations ($b = 0.10$, $p = .05$). For non–African American students the path from expectations and responsiveness to developmental needs to post mathematics achievement was not statistically significantly different from zero ($b = -0.08$, $p = .57$) and there was no mediation ($b = -0.01$, $p = .33$, 95% CI [−.05, .02]). For African American students, expectations and responsiveness to developmental needs was significantly associated with post mathematics achievement ($b = 0.54$, $p < .001$). However, the parametric bootstrap test for mediation only approached statistical significance ($b = 0.05$, $p = .07$, 95% CI [0.002, 0.12]). These findings suggest that African American students especially benefited from teacher expectations and responsiveness to developmental needs.

A Wald test comparing the point estimates of the path from expectations and responsiveness to developmental needs of the two groups (African American and non–African American students) was statistically significant ($p < .001$). The coefficients, standard errors, *p* values, and 95% confidence intervals for the full list of variables are displayed in Table 3.

In the model comparing White and Hispanic students (488 students in 88 classrooms), the treatment did not quite have a significant impact on teacher expectations ($b = .09$ $p = .06$). For Hispanic students, the path from expectations and responsiveness to developmental needs to post mathematics achievement was not significantly different from zero ($b = -0.21$, $p = .33$). For White students, expectations and responsiveness to developmental needs was not significantly associated with post mathematics achievement ($b = 0.03$, $p = .93$). The parametric bootstrap test for mediation was not statistically significant for either group. The full table (Table S1) of results is displayed in the supplemental materials.

Mediator 2: Teacher Confidence and Enthusiasm

Figure 2 displays the path diagram for the second mediator: teacher confidence and enthusiasm. Teachers in the treatment group on average were rated 0.25 points higher ($p = .001$) on teacher confidence and enthusiasm than teachers in the control group.

For non–African American students, the path from teacher confidence and enthusiasm to post mathematics achievement was not significantly different from zero ($b = 0.14$, $p = .12$). The parametric bootstrap test indicated no statistically significant mediation for non–African American students ($b = 0.04$, $p = .10$, 95% CI [0.0003, 0.08]). For African American students, teacher confidence and enthusiasm was statistically significantly associated with posttreatment mathematics achievement ($b = 0.25$, $p = .001$) such that a one-unit increase in observed teacher confidence and enthusiasm yielded a predicted 0.25 unit increase in their posttreatment mathematics scores. The parametric bootstrap test indicated partial mediation for African American students ($b = 0.06$, $p = .03$, 95% CI [0.01, 0.13]). This implies that part of the effects of the treatment were through increasing teacher's confidence and enthusiasm for teaching.

Table 2
Correlations Among Classroom Observations, Teacher Characteristics, and Students' Mathematics Achievement in 101 Classrooms (N = 1,238)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. T1 math	1.00																			
2. T2 math	**.57**	1.00																		
3. Male	−.06	−.06	1.00																	
4. African American	**−.11**	**−.20**	.02	1.00																
5. Hispanic	**−.10**	.03	−.03	**−.58**	1.00															
6. LEP	**−.14**	.01	−.01	**−.40**	**.50**	1.00														
7. Special education	−.03	−.04	.01	−.07	.07	−.01	1.00													
8. Age	**.29**	**.30**	.05	**−.20**	**.25**	**.24**	**.08**	1.00												
9. L2 confidence	**.09**	**.19**	−.01	**.08**	**−.12**	**−.18**	.03	**−.09**	1.00											
10. L2 expectations	**.08**	**.09**	−.04	**.12**	**−.28**	**−.29**	.01	**−.19**	**.60**	1.00										
11. L2 discourse	.02	−.03	−.03	**.20**	**−.36**	**−.36**	−.05	**−.30**	**.43**	**.62**	1.00									
12. L2 number of computers	.01	**.18**	.02	**−.10**	**.08**	**.09**	.02	−.03	**.22**	**.14**	−.02	1.00								
13. L2 number of math activities	.05	**.27**	−.01	**−.08**	**.13**	**.11**	.05	**.17**	**.31**	**.11**	**−.26**	**.44**	1.00							
14. L2 minutes of math	.06	**.11**	.02	.03	**−.10**	**−.08**	−.02	.02	**.24**	**.12**	−.02	**.14**	**.38**	1.00						
15. L2 assign math homework	−.01	**.10**	−.05	.03	**.11**	**.12**	.05	**.10**	**−.08**	**−.10**	**−.28**	**.19**	**.35**	.01	1.00					
16. L2 classroom size	.02	**.10**	−.01	**−.13**	**.22**	**.24**	.05	**.22**	.06	**−.10**	**−.19**	**.10**	**.31**	.02	**.21**	1.00				
17. L2 Master's degree	.00	−.02	−.00	**.17**	**−.33**	**−.24**	−.04	**−.23**	**.18**	**.29**	**.44**	.02	**−.19**	**.13**	−.05	**−.27**	1.00			
18. L2 No. years teaching	−.06	−.04	.02	**.11**	−.03	**−.08**	−.02	**−.11**	**−.09**	**−.09**	−.05	.03	−.07	−.05	−.04	**−.20**	**.08**	1.00		
19. L2 PD hours	−.06	−.01	.03	**.13**	**−.08**	**−.08**	−.01	**−.11**	**−.08**	**−.08**	−.03	.07	−.03	−.04	.00	**−.17**	**.20**	**.96**	1.00	
20. L2 African American	**−.08**	**−.13**	.04	**.69**	**−.49**	**−.43**	−.06	**−.22**	**.12**	**.18**	**.29**	**−.14**	**−.12**	.05	.04	**−.18**	**.25**	**.17**	**.19**	1.00
21. Percent SES	**.14**	**.13**	−.03	−.01	**−.09**	**−.18**	.04	**−.11**	**.23**	**.14**	.05	**.25**	**.08**	−.01	**.11**	.00	**.17**	.00	.07	−.01

Note. LEP = limited English proficient; PD = professional development; SES = socioeconomic status. Bolded correlations are statistically significant at $p < .01$. L2 is classroom level.

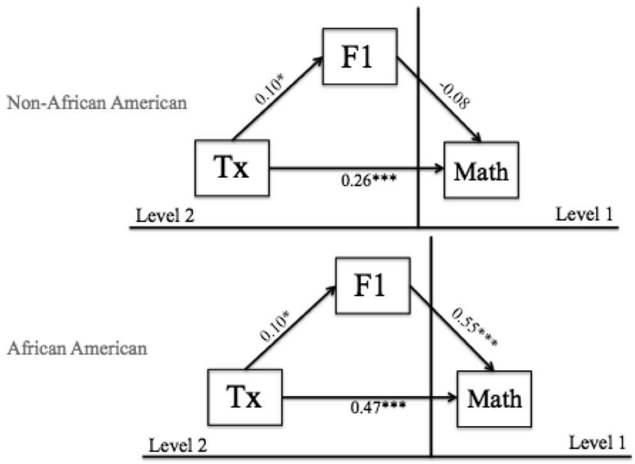

Figure 1. Multigroup multilevel mediation for expectations and responsiveness to developmental needs. Unstandardized regression coefficients at the between-level are shown. Wald test was statistically significant. Results from the parametric bootstrap test for mediation for the African American group suggest partial mediation approaching statistical significance. Tx = treatment; F1 = expectations and responsiveness to developmental needs. $^{*}p < 0.05$. $^{**}p < 0.01$. $^{***}p < 0.001$.

Although we found evidence of partial mediation for African American students but not for non–African American students, a Wald test comparing the point estimates of the path from teacher confidence and enthusiasm to post mathematics achievement was not statistically significant (p = .28). It should be noted that the confidence interval for the point estimate for the non–African American group includes the point estimate from the African American group suggesting that the effects for the two groups might not actually differ. As such, we suggest that limited evi-

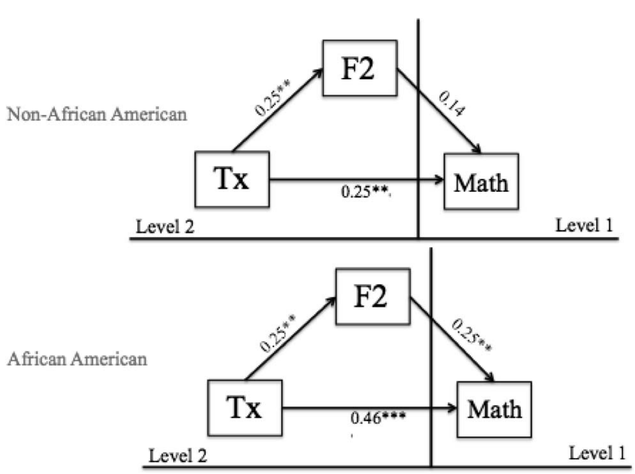

Figure 2. Multigroup multilevel mediation for teacher confidence and enthusiasm. Unstandardized regression coefficients at the between-level are shown. Wald test was not statistically significant. Results from the parametric bootstrap test for mediation suggest partial mediation for African American students but no mediation for non–African American students. Tx = treatment; F2 = teacher confidence and enthusiasm. $^{*}p <$ 0.05. $^{**}p < 0.01$. $^{***}p < 0.001$.

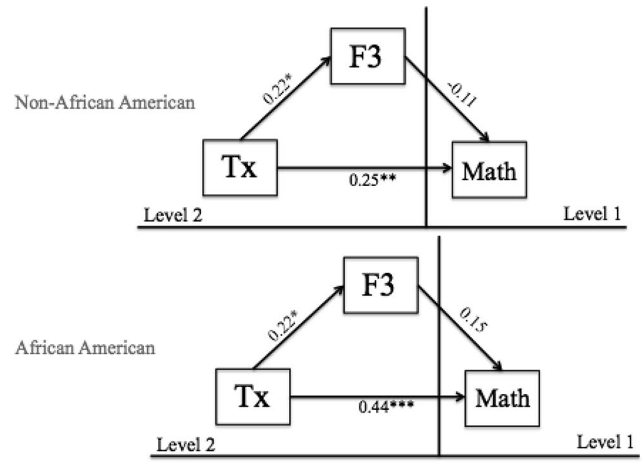

Figure 3. Multigroup multilevel mediation for support for mathematical discourse. Unstandardized regression coefficients at the between-level are shown. No significant mediation. Tx = treatment; F3 = support for mathematical discourse. $^{*}p < 0.05$. $^{**}p < 0.01$. $^{***}p < 0.001$.

dence supports the conclusion that teacher confidence and enthusiasm differentially influenced African American students' mathematics achievement. The coefficients, standard errors, p values, and 95% confidence intervals for the full list of variables are displayed in Table 4.

In the model comparing White and Hispanic students (488 students in 88 classrooms), the treatment had a positive impact on teacher confidence (b = .20, p = .02). For Hispanic students the path from teacher confidence to post mathematics achievement was not statistically significantly different from zero (b = 0.06, p = .68). For White students, teacher confidence was also not statistically significantly associated with post mathematics achievement (b = 0.19, p = .12). The parametric bootstrap test was not statistically significant for either group. The full table (Table S2) of results is displayed in the supplemental materials.

Mediator 3: Support for Mathematical Discourse

Figure 3 displays the path diagram for the third mediator: support for mathematical discourse. Teachers in the treatment group on average were rated 0.22 points higher (p = .01) on support for mathematical discourse than teachers in the control group.

The path from support for mathematical discourse to posttreatment mathematics achievement was not statistically significant for both groups (b = −0.11, p = .24; b = 0.16, p = .13, for non–African American and African American students, respectively). As such, the parametric bootstrap test was not statistically significant for either group (b = −0.02, p = .46, 95% CI [−0.08, 0.02]; b = 0.03, p = .22, 95% CI [−0.01, 0.10], for non–African American and African American students, respectively). The coefficients, standard errors, p values, and 95% confidence intervals for the full list of variables are displayed in Table 5.

In the model comparing White and Hispanic students (488 students in 88 classrooms), the treatment had a positive impact on support for mathematical discourse (b = .21, p = .02). For Hispanic students, the path from mathematical discourse to post

Table 3
Results From the Multigroup Multilevel Mediation Model With Expectations and Responsiveness to Developmental Needs as the Mediator (N = 1,238 Students in 101 Classrooms)

Mediator	African American students			Non–African American students		
	β (*SE*)	95% CI	*p*	β (*SE*)	95% CI	*p*
Indirect effect (Treatment → Post-mathematics)						
Within level						
Pre-mathematics	.44 (.03)	[.40–.49]	<.001	.44 (.03)	[.40–.49]	<.001
Male	−.05 (.03)	[−.10–.00]	.12	−.05 (.03)	[−.10–.00]	.12
Age	.26 (.05)	[.18–.35]	<.001	.26 (.05)	[.18–.35]	<.001
LEP	.07 (.05)	[−.01–.16]	.14	.07 (.05)	[−.01–.16]	.14
Special education	−.07 (.04)	[−.14–−.01]	.07	−.07 (.04)	[−.14–−.01]	.07
Hispanic	—	—	—	−.11 (.05)	[−.20–−.03]	.03
Between level						
Treatment	.47 (.07)	[.37–.58]	<.001	.26 (.07)	[.14–.37]	<.001
Treatment → Expectations and responsiveness to developmental needs						
Between level						
Treatment	.10 (.05)	[.02–.18]	.05	.10 (.05)	[.02–.18]	.05
Expectations and responsiveness to developmental needs → Post-mathematics						
Between Level						
Expectations	.55 (.13)	[.33–.77]	<.001	−.08 (.15)	[−.33–.16]	.57
Number of computers	.00 (.02)	[−.03–.03]	.98	.00 (.02)	[−.03–.03]	.98
Number of math activities	.03 (.01)	[.01 −.05]	.003	.03 (.01)	[.01–.05]	.003
Minutes of math	−.001 (.002)	[−.003–.002]	.69	−.001 (.002)	[−.003–.002]	.69
Assign math homework	.02 (.02)	[−.03–.02]	.54	.02 (.02)	[−.03–.02]	.54
Classroom size	−.01 (.01)	[−.02–.01]	.31	−.01 (.01)	[−.02–.01]	.31
Master's degree	−.03 (.07)	[−.13–.08]	.7	−.03 (.07)	[−.13–.08]	.7
Number of years teaching	−.01 (.01)	[−.02–.01]	.41	−.01 (.01)	[−.02–.01]	.41
PD hours	.01 (.01)	[−.01–.02]	.35	.01 (.01)	[−.01–.02]	.35
Percent African American	.06 (.09)	[−.11–.15]	.50	.06 (.09)	[−.11–.15]	.50
Percent SES	−.01 (.12)	[−.20–.19]	.64	−.01 (.12)	[−.20–.19]	.64

Note. Unstandardized coefficients are reported. PD = professional development; SES = socioeconomic status; *SE* = standard error; CI = confidence interval; LEP = limited English proficient. Expectations is expectations and responsiveness to developmental needs. PD hours is number of PD hours before the intervention.

mathematics achievement was not statistically different from zero ($b = -0.22$, $p = .11$). For White students, support for mathematical discourse was also not significantly associated with post mathematics achievement ($b = -0.02$, $p = .86$). The parametric bootstrap test for mediation was not statistically significant for both groups. The full table (Table S3) of results is displayed in the supplemental materials.

Discussion

This study evaluated the differential effects of three instructional practices on African American students' achievement in the context of a randomized control trial of preschool classrooms. Specifically, we found that, on average, African American students benefited more from certain instructional practices (teacher expectations and responsiveness to developmental needs) than non–African American students and that teacher expectations and responsiveness to developmental needs only approached significance for the partially mediated effect of the intervention on African American students' mathematics achievement. No support was found for the effect of support for mathematical discourse on student achievement for either group.

In this study, we compared African American students with non–African American students such that White *and* Hispanic students were included in the former category. This decision was further justified when we replicated our analyses and restricted the sample to just those two groups and found no statistically significant differences between the groups. In recent years, Hispanic and African American children have shown different achievement trajectories relative to White students (National Center for Education Statistics, 2013; Reardon, Robinson-Cimpian, & Weathers, 2015; for an exception see Rumberger & Palardy, 2005). Whereas White-African American achievement gaps increase during the first six years of schooling in both math and reading, White-Hispanic gaps decrease during this period (Fryer & Levitt, 2004, 2006; Hemphill & Vanneman, 2011; Reardon & Galindo, 2006; Reardon & Robinson, 2008). Our findings are consistent with Bodovski and Farkas (2007) and Bottia and colleagues (2014) who reported modest reductions in the achievement gap between African American and White students in kindergarten as a result of certain instructional practices, but no reductions in the achievement gap between Whites and Hispanics. For example, instruction that focuses on developmentally appropriate but higher-level thinking has been shown to have a positive impact on African American but not Hispanic children's mathematics performance (Bodovski & Farkas, 2007). It is unclear from the empirical and theoretical literature why African American and Hispanic children responded differentially to the classroom environment. Theory

Table 4
Results From the Multigroup Multilevel Mediation Model With Teacher Confidence and Enthusiasm as the Mediator (N = 1,238 Students in 101 Classrooms)

	African American students			Non–African American students		
Mediator	β (*SE*)	95% CI	*p*	β (*SE*)	95% CI	*p*
Indirect effect (Treatment → Post-mathematics)						
Within level						
Pre-mathematics	.44 (.02)	[.41–.48]	<.001	.44 (.02)	[.41–.48]	<.001
Male	−.05 (.03)	[−.10–.01]	.07	−.05 (.03)	[−.10–.01]	.07
Age	.27 (.05)	[.18–.35]	<.001	.27 (.05)	[.18–.35]	<.001
LEP	.08 (.05)	[−.002–.17]	.11	.08 (.05)	[−.002–.17]	.11
Special education	−.08 (.04)	[−.14–−.01]	.05	−.08 (.04)	[−.14–−.01]	.05
Hispanic	—	—	—	−.10 (.05)	[−.19–−.02]	.05
Between level						
Treatment	.46 (.07)	[.35–.57]	<.001	.25 (.07)	[.13–.37]	.001
Treatment → Teacher confidence and enthusiasm						
Between level						
Treatment	.25 (.08)	[.13–.37]	.001	.25 (.08)	[.13–.37]	.001
Teacher confidence and enthusiasm → Post-mathematics						
Between level						
Confidence	.25 (.08)	[.13–.38]	.001	.14 (.07)	[−.01–.29]	.03
Number of computers	−.003 (.02)	[−.03–.03]	.89	−.003 (.02)	[−.03–.03]	.89
Number of math activities	.02 (.01)	[.002−.04]	.07	.02 (.01)	[.002− .04]	.07
Minutes of math	−.00 (.002)	[−.003–.002]	.79	−.00 (.002)	[−.003–.002]	.79
Assign math homework	.03 (.02)	[−.01–.07]	.23	.03 (.02)	[−.01–.07]	.23
Classroom size	−.01 (.01)	[−.02–.003]	.23	−.01 (.01)	[−.02–.003]	.23
Master's degree	−.06 (.07)	[−.17–.05]	.39	−.06 (.07)	[−.17–.05]	.39
Number of years teaching	−.01 (.01)	[−.02–.01]	.30	−.01 (.01)	[−.02–.01]	.3
PD hours	.01 (.01)	[−.003–.02]	.21	.01 (.01)	[−.003–.02]	.21
Percent African American	.02 (.08)	[−.11–.15]	.77	.02 (.08)	[−.11–.15]	.77
Percent SES	−.01 (.12)	[−.20–.19]	.96	−.01 (.12)	[−.20–.19]	.96

Note. Unstandardized coefficients are reported. PD = professional development; SES = socioeconomic status; *SE* = standard error; CI = Confidence Interval; LEP = limited English proficient. Confidence is teacher confidence and enthusiasm. PD hours is number of PD hours before the intervention.

from García Coll and colleagues (1996) suggests that children of different racial/ethnic groups hold differing cultural values, experiences of prejudice, and family values, which may all be factors in children's perceptions of and participation in the learning environment. This issue is still not well understood and future research is needed to better understand how racial/ethnic background influences students' experiences of the learning environment. It should be noted that in the present study, we only examined three teacher practices and it is quite possible that an examination of other classroom dimensions would yield instances where Hispanic students respond differentially to the classroom environment than White students.

A contribution of this study is that we address a limitation of previous studies investigating racial/ethnic differences in responses to the classroom (e.g., Bodovski & Farkas, 2007; Bottia et al., 2014; Sonnenschein & Galindo, 2015). Specifically, we were able to directly investigate within-classroom variation in responsiveness to instructional practices rather than relying on a dataset that was collected at the student-level. A student-level dataset is problematic because observations or teacher reports of the *same* instructional practices were not collected and it could not be investigated whether students, within the same classroom, differentially benefit from certain instructional practices. It may suggest that other confounding factors such as access to quality classrooms or the actual classroom environment children experience may influence the interpretation of previous studies. Additionally, instructional practices in previous studies (e.g., Bodovski & Farkas, 2007; Bottia et al., 2014; Sonnenschein & Galindo, 2015) were measured using teacher reports, which may have limited reliability.

Teacher Expectations and Developmental Appropriateness

Findings from our study extend the research on teacher expectations and student achievement in the following ways: (a) observed teacher expectations can be modified through professional development implemented as part of a mathematics intervention, (b) teacher expectations differentially influence mathematics achievement for African American students, and (c) teacher expectations significantly influence students' mathematics achievement in young students (mean age 4.34 years) supporting earlier work suggesting the importance of early expectations (Alvidrez & Weinstein, 1999). In support of theory on teacher expectancy beliefs, the effects we found in our sample of preschool students may have been strongest during that year because teachers had little prior contact with children, the teacher was integral in conveying the mathematics knowledge to students, and the mathematics content was novel (West & Anderson, 1976). It may be that African American students respond positively to these kinds of teacher practices and behaviors because these practices and beliefs have been found to be absent from environments (school, home, or otherwise) where large numbers of African Americans participate (Oakes, 1990). Theoretical work on African American students'

Table 5
Results From the Multigroup Multilevel Mediation Model With Support for Mathematical Discourse as the Mediator (N = 1,238 Students in 101 Classrooms)

	African American students			Non–African American students		
Mediator	β (*SE*)	95% CI	*p*	β (*SE*)	95% CI	*p*
Indirect effect (Treatment → Post-mathematics)						
Within level						
Pre-mathematics	.45 (.03)	[.40−.51]	<.001	.45 (.03)	[.40−.51]	<.001
Male	−.05 (.03)	[−.10−.003]	.13	−.05 (.03)	[−.10−.003]	.13
Age	.26 (.05)	[.17−.34]	<.001	.26 (.05)	[.17−.34]	<.001
LEP	.07 (.05)	[−.02−.15]	.2	.07 (.05)	[−.02−.15]	.20
Special education	−.07 (.04)	[−.14−−.01]	.07	−.07 (.04)	[−.14−−.01]	.07
Hispanic	—	—	—	−.12 (.05)	[−.21−−.04]	.02
Between level						
Treatment	.44 (.08)	[.31−.57]	<.001	.25 (.08)	[.12−.38]	.001
Treatment → Support for mathematical discourse						
Between level						
Treatment	.22 (.09)	[.08–.36]	.01	.22 (.09)	[.08–.36]	.01
Support for mathematical discourse → Post-mathematics						
Between level						
Discourse	.15 (.10)	[−.01–.33]	.13	−.11 (.10)	[−.27–.05]	.24
Number of computers	.01 (.02)	[−.03–.04]	.80	.01 (.02)	[−.03–.04]	.80
Number of math activities	.03 (.01)	[.02–.05]	.001	.03 (.01)	[.02–.05]	.001
Minutes of math	.00 (.002)	[−.003–.002]	.83	.00 (.002)	[−.003–.002]	.83
Assign math homework	.01 (.03)	[−.03–.05]	.70	.01 (.03)	[−.03–.05]	.70
Classroom size	−.01 (.01)	[−.02–.004]	.27	−.01 (.01)	[−.02–.004]	.27
Master's degree	−.02 (.07)	[−.14–.10]	.80	−.02 (.07)	[−.14–.10]	.80
Number of years teaching	−.01 (.01)	[−.02–.01]	.44	−.01 (.01)	[−.02–.01]	.44
PD hours	.01 (.01)	[−.01–.02]	.37	.01 (.01)	[−.01–.02]	.37
Percent African American	.05 (.09)	[−.10–.19]	.61	.05 (.09)	[−.10–.19]	.61
Percent SES	.07 (.14)	[−.15–.29]	.60	.07 (.14)	[−.15–.29]	.60

Note. Unstandardized coefficients are reported. PD = professional development; SES = socioeconomic status; *SE* = standard error; CI = Confidence Interval; LEP = limited English proficient. Discourse is support for mathematical discourse. PD hours is number of PD hours before the intervention.

mathematics learning also suggests the expectations teachers have for these students are important (Ladson-Billings, 1997). It may also be that African American students are typically not exposed to environments where individuals have high expectations of them (Ferguson, 2000; Lopez, 2002). However, most of the research suggesting teacher expectations are especially important for African American students is usually qualitative in nature and does not compare the effects of teacher expectations on achievement across different racial/ethnic groups. The present study provided empirical support for this assertion but cannot answer why. More research using student and teacher interviews could shed light on this question.

The expectations teachers have of their students may be an indicator of overall classroom quality. For example, Rubie-Davies (2007) found statistically significant differences in observed practices of teachers who were low, average, and high in their expectations of students. Teachers in the high expectations group were observed to provide more feedback to their students, engage in more higher-order questioning, and were rated higher in managing student behavior. In the present study, we found evidence of this in the large and statistically significant correlations among the three classroom dimensions we measured. These findings suggest that efforts to improve teachers' instructional practices could also include professional development opportunities aimed at changing teacher's expectations of their students and can even be implemented at the same time as interventions focused on improving mathematics achievement. This integrated view of mathematics whereby improvement of elements of mathematics instruction *and* the classroom climate is supported in the NCTM standards (NCTM, 2000). Additionally, we note that it is important to consider multiple dimensions of the classroom-learning environment as only looking at teacher expectancy effects could mask our understanding of other important dimensions.

Teacher Confidence and Enthusiasm

The intervention significantly increased the observed teacher confidence and enthusiasm of teachers in the treatment condition. This is not surprising given that the intervention was mathematics-specific and provided treatment teachers with ample support to teach mathematics through the *Building Blocks* curriculum. Additionally, we found some support that teacher confidence and enthusiasm differentially partially mediated the effect of the intervention on African American students' mathematics achievement but not non–African American students' achievement. Even though we found a statistically significant association between observer's ratings of teacher confidence and enthusiasm and African American students' mathematics achievement, a Wald test comparing the coefficients between the two groups was not statistically significant, suggesting that there may not actually be a statistically significant association between teacher confidence and enthusiasm and African American students' mathematics achievement. We caution readers from over interpreting these findings and instead suggest that future research replicate this finding. Previous research does, however, suggest that teachers' mathematical knowledge for teaching is associated with gains in student achievement over time (Hill, Rowan, & Ball, 2005) and that mathematical knowledge for teaching could be related to how confidently and enthusiastically a teacher teaches.

Support for Mathematical Discourse

Although we found that the intervention significantly increased teacher's observed support for mathematical discourse, we did not find that this dimension was significantly associated with students' posttreatment mathematics achievement. This was surprising and stands in contrast to theory suggesting the importance of support for mathematical discourse and students' mathematics achievement (e.g., Chi, 2000). In the present study, it may be that support for the mathematical discourse dimension did not have an additional effect on students' achievement above and beyond the elements that were already present in the intervention, such as the use and quality of the *Building Blocks* software. Additionally, the age of the students in our sample may be too young to benefit from high quality support for mathematical discourse. Because observations were conducted of the *teacher's* support for mathematical discourse, it is unclear if the preschool students in our sample actually took up this practice and benefited from it.

Limitations and Future Directions

We note several limitations of this work. Unfortunately, we did not have information about the teacher's ethnicity and therefore could not investigate effects based on teacher and student ethnicity match. Conflicting evidence for the importance of teacher–child ethnicity match exists on outcomes such as achievement and school behavioral adjustment. For example, Ewing and Taylor (2009) did not find that teacher–child ethnicity match moderated the association between relationship quality and behavioral adjustment. However, Saft and Pianta (2001) found that teacher–child ethnicity match was related to teacher's expectations of the child. It was not possible for us to examine teacher and student ethnicity match with our data, but it should be a direction for future research.

The design of the present study prevents us from teasing apart what particular aspects of the intervention led to the differential improvements in student performance. It remains unclear whether this might be attributable to the additional support teachers received during the professional development sessions or the novel attention the research team paid to the teachers. A useful direction for future research would be to test the relative effectiveness of each of the intervention components separately in the form of a treatment "dismantling approach" in which only certain components of mathematics interventions already known to be effective are retained. Further, it is important to note that a unique advantage of our study is that we were able to show that teachers' instructional behaviors, such as holding high expectations for all students, can be changed within the context of a mathematics intervention. Future studies should also consider including a second type of intervention, such as a literacy or socioemotional curriculum, to compare the effects of teachers' general participation in the intervention to the *Building Blocks* intervention.

In considering why many of the proposed classroom mediators were nonsignificant, it could be that other unmeasured characteristics of the classroom were stronger mediators of the intervention for African American students. Another speculation is that whereas classroom environments help support the learning and achievement of students, it is really students' participation in those environments that influences their subsequent achievement (e.g., Ing et al., 2014; Ruzek et al., 2016; Skinner & Belmont, 1993). Ing and colleagues (2014), for example, coded video observations of elementary school mathematics classrooms for the quality of students' engagement the classroom. The authors coded for the level of detail students gave in their explanations as well as the quality of students' engagement and found that the quality of students' engagement related to gains in standardized mathematics achievement. Though these analyses are time and researcher intensive, future research should consider incorporating measures of student engagement—whether through observations or self-report—in understanding the relation between instructional characteristics and student achievement.

We examined the importance of classroom learning environments for students' development of mathematics skills; however, we were unable to examine whether sustained high-quality instruction, and in the case of the present study, sustained teacher expectations could aid in closing the achievement gap between African American students and other groups. A possible fruitful area of future research could be to examine the same associations we investigated in this paper but include more information on the learning environments students are subsequently exposed to. Looking at sustained high-quality instruction could help us understand whether these differential effects of the classroom environment remain even in cases where African American students have had teachers with high levels of confidence and enthusiasm or high expectations for them. Earlier work documenting the middle school and high school transition (Hirsch & Rapkin, 1987; Midgley, Feldlaufer, & Eccles, 1989; Wigfield, Eccles, Mac Iver, Reuman, & Midgley, 1991) has found that continuity in students' experiences mattered for their outcomes. For example, Midgley and colleagues (1989) looked at students' transition from elementary school to junior high and found that students who had rated their teacher as low on support in elementary school but went into classrooms where they rated their teacher as high in support experienced higher intrinsic motivation than students who had initially highly supportive teachers and transitioned into classrooms where they were low in support. This would also help us understand possible reasons for the fadeout effect in educational interventions. Will students who are continually in high-quality classrooms sustain gains that are not found in interventions that only last one year? Rubie-Davies and colleagues (2014) investigated the cumulative effects of teachers from kindergarten to fourth grade on students' fourth grade academic achievement and found that the higher the teacher's expectations were relative to the student's actual ability, the higher the student's achievement in fourth grade. Conversely, the lower the teachers' expectations were over time, the lower the student achieved at the end of fourth grade. In understanding students' continuity of classroom experiences, information on the cumulative effects of classroom environments could be better explored.

Conclusion

Taken together, our findings suggest that specific dimensions of the classroom environment differentially influenced African American students' mathematics learning. Specifically, teacher expectations and developmental appropriateness increased the achievement of African

American students. Additionally, we found that a carefully designed and well-implemented preschool mathematics intervention changed multiple dimensions of the classroom environment. If classroom interventions are likely to help close important race-based achievement gaps, then targeting classroom processes, such as teacher expectations, might prove fruitful in making progress toward this goal. Findings from this study can help researchers and practitioners understand the ways in which the organization of classroom learning environments might be structured more suitably to leverage and support the positive mathematics learning experiences of students who stand to benefit from it the most.

References

Abbott-Shim, M., Lambert, R., & McCarty, F. (2000). Structural model of Head Start classroom quality. *Early Childhood Research Quarterly, 15,* 115–134. http://dx.doi.org/10.1016/S0885-2006(99)00037-X

Aikens, N. L., & Barbarin, O. (2008). Socioeconomic differences in reading trajectories: The contribution of family, neighborhood, and school contexts. *Journal of Educational Psychology, 100,* 235–251. http://dx.doi.org/10.1037/0022-0663.100.2.235

Alexander, K. L., Entwisle, D. R., & Horsey, C. S. (1997). From first grade forward: Early foundations of high school dropout. *Sociology of Education, 70,* 87–107. http://dx.doi.org/10.2307/2673158

Alvidrez, J., & Weinstein, R. S. (1999). Early teacher perceptions and later student academic achievement. *Journal of Educational Psychology, 91,* 731–746. http://dx.doi.org/10.1037/0022-0663.91.4.731

Askew, M., Brown, M., Rhodes, V., Wiliam, D., & Johnson, D. (1997). *Effective teachers of numeracy: Report of a study carried out for the Teacher Training Agency*. London, UK: King's College, University of London.

Asparouhov, T., & Muthén, B. O. (2012). *Multiple group multilevel analysis* (Mplus Web Notes, No. 16). Retrieved from www.statmodel.com/examples/webnotes/webnote16.pdf

Battey, D. (2013). "Good" mathematics teaching for students of color and those in poverty: The importance of relational interactions within instruction. *Educational Studies in Mathematics, 82,* 125–144. http://dx.doi.org/10.1007/s10649-012-9412-z

Berry, R. Q., III. (2003). Mathematics standards, cultural styles, and learning preferences: The plight and the promise of African American students. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 76,* 244–249. http://dx.doi.org/10.1080/00098650309602013

Bishop-Josef, S. J., & Zigler, E. (2011). The cognitive/academic emphasis versus whole child approach: The 50-year debate. In E. Zigler, W. S. Gilliam, & W. S. Barnett (Eds.), *The pre-K debates. Current controversies and issues* (pp. 83–88). Baltimore, MD: Brookes

Boaler, J. (1998). Open and closed mathematics: Student experiences and understandings. *Journal for Research in Mathematics Education, 29,* 41–62. http://dx.doi.org/10.2307/749717

Bodovski, K., & Farkas, G. (2007). Do instructional practices contribute to inequality in achievement? The case of mathematics instruction in kindergarten. *Journal of Early Childhood Research, 5,* 301–322. http://dx.doi.org/10.1177/1476718X07080476

Bottia, M. C., Moller, S., Mickelson, R. A., & Stearns, E. (2014). Foundations of mathematics achievement. *The Elementary School Journal, 115,* 124–150. http://dx.doi.org/10.1086/676950

Brooks-Gunn, J., & Duncan, G. J. (1997). The effects of poverty on children. *Future Child, 7,* 55–71.

Brophy, J. (1986). Teacher influences on student achievement. *American Psychologist, 41,* 1069–1077. http://dx.doi.org/10.1037/0003-066X.41.10.1069

Brophy, J. E., & Good, T. L. (1970). Teachers' communication of differential expectations for children's classroom performance: Some behavioral data. *Journal of Educational Psychology, 61,* 365–374. http://dx.doi.org/10.1037/h0029908

Bryant, D. M., Burchinal, M., Lau, L. B., & Sparling, J. J. (1994). Family and classroom correlates of Head Start children's developmental outcomes. *Early Childhood Research Quarterly, 9*(3–4), 289–309. http://dx.doi.org/10.1016/0885-2006(94)90011-6

Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology: Educational design and cognitive science* (pp. 161–238). Hillsdale, NJ: Erlbaum.

Clarke, J., Frazer, E., DiMartino, J., Fisher, P., & Smith, P. (2003). Making learning personal: Educational practices that work. In J. DiMartino, J. Clarke, & D. Wolk (Eds.) *Personalized learning: Preparing high school students to create their futures* (pp. 173–194). Lanham, MD: Scarecrow Press.

Clements, D. H., Baroody, A. J., & Sarama, J. (2013). Background research on early mathematics. *National Governor's Association, Center Project on Early Mathematics*.

Clements, D. H., & Sarama, J. (2016). *COEMET: The Classroom Observation of Early Mathematics Environment and Teaching instrument*. Denver, CO: University of Denver. (Original work published 2000)

Clements, D. H., & Sarama, J. (2007). Effects of a preschool mathematics curriculumml: Summative research on the Building Blocks project. *Journal for Research in Mathematics Education, 38,* 136–163.

Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal, 45,* 443–494. http://dx.doi.org/10.3102/0002831207312908

Clements, D. H., Sarama, J. H., & Liu, X. H. (2008). Development of a measure of early mathematics achievement using the Rasch model: The Research-Based Early Maths Assessment. *Educational Psychology, 28,* 457–482. http://dx.doi.org/10.1080/01443410701777272

Clements, D. H., Sarama, J., Spitler, M. E., Lange, A. A., & Wolfe, C. B. (2011). Mathematics learned by young children in an intervention based on learning trajectories: A large-scale cluster randomized trial. *Journal for Research in Mathematics Education, 42,* 127–166.

Clements, D. H., Sarama, J., Wolfe, C. B., & Spitler, M. E. (2013). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies Persistence of effects in the third year. *American Educational Research Journal, 50,* 812–850. http://dx.doi.org/10.3102/0002831212469270

Darling-Hammond, L. (1997). *The right to learn: A blueprint for creating schools that work*. San Francisco, CA: Jossey-Bass.

Diaz, R. M. (2008). *The role of language in early childhood mathematics* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database (UMI No. 3319002).

Duncan, G. J., Brooks-Gunn, J., & Klebanov, P. K. (1994). Economic deprivation and early childhood development. *Child Development, 65,* 296–318.

Duncan, G., & Magnuson, K. (2005). Can family socioeconomic resources account for racial and ethnic test score gaps? *The Future of Children, 15,* 35–54. http://dx.doi.org/10.1353/foc.2005.0004

Dunn, L., & Kontos, S. (1997). What have we learned about developmentally appropriate practice? Research in review. *Young Children, 52,* 4–13.

Dusek, J. B., & Joseph, G. (1983). The bases of teacher expectancies: A meta-analysis. *Journal of Educational Psychology, 75,* 327–346. http://dx.doi.org/10.1037/0022-0663.75.3.327

Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science, 1,* 54–75. http://dx.doi.org/10.1214/ss/1177013815

Evertson, C. M., Emmer, E. T., & Brophy, J. E. (1980). Predictors of effective teaching in junior high mathematics classrooms. *Journal for Research in Mathematics Education, 11,* 167–178. http://dx.doi.org/10.2307/748938

Ewing, A. R., & Taylor, A. R. (2009). The role of child gender and ethnicity in teacher–child relationship quality and children's behavioral adjustment in preschool. *Early Childhood Research Quarterly, 24,* 92–105. http://dx.doi.org/10.1016/j.ecresq.2008.09.002

Ferguson, A. A. (2000). *Bad boys: Public schools in the making of black masculinity*. Ann Arbor, MI: University of Michigan Press. http://dx.doi.org/10.3998/mpub.16801

Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation, 28,* 465–498.

Franke, M. L., Fennema, E., & Carpenter, T. (1997). Teachers creating change: Examining evolving beliefs and classroom practice. *Mathematics Teachers in Transition,* 255–282.

Fryer, R. G., Jr., & Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. *The Review of Economics and Statistics, 86,* 447–464. http://dx.doi.org/10.1162/003465304323031049

Fryer, R. G., & Levitt, S. D. (2006). The black-white test score gap through third grade. *American Law and Economics Review, 8,* 249–281. http://dx.doi.org/10.1093/aler/ahl003

García Coll, C., Lamberty, G., Jenkins, R., McAdoo, H. P., Crnic, K., Wasik, B. H., & Vázquez García, H. (1996). An integrative model for the study of developmental competencies in minority children. *Child Development, 67,* 1891–1914. http://dx.doi.org/10.2307/1131600

Gill, S., & Reynolds, A. J. (1999). Educational expectations and school achievement of urban African American children. *Journal of School Psychology, 37,* 403–424. http://dx.doi.org/10.1016/S0022-4405(99)00027-8

Ginsburg, H. P., Inoue, N., & Seo, K. (1999). Young children doing mathematics: Observations of everyday mathematics. In J. Copley (Ed.), *Mathematics in the early years* (pp. 88–100). Washington, DC: National Association for the Education of Young Children.

Ginsburg, H., Lee, J., & Boyd, J. (2008). Mathematics education for young children: What it is and how to promote it. *Social Policy Report,* XX11, 1. Retrieved from http://www.scrd.org

Haberman, M. (1991). The pedagogy of poverty versus good teaching. *Phi Delta Kappan, 73,* 290–294.

Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early childhood environment rating scale* (rev. ed.). New York, NY: Teachers College Press.

Hemphill, F. C., & Vanneman, A. (2011). *Achievement gaps: How Hispanic and white students in public schools perform in mathematics and reading on the National Assessment of Educational Progress* (NCES 2011–459). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Henningsen, M., & Stein, M. K. (1997). Mathematical tasks and student cognition: Classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. *Journal for Research in Mathematics Education, 28,* 524–549. http://dx.doi.org/10.2307/749690

Hiebert, J. C., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester, Jr., (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 1, pp. 371–404). New York, NY: Information Age Publishing.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42,* 371–406. http://dx.doi.org/10.3102/00028312042002371

Hirsch, B. J., & Rapkin, B. D. (1987). The transition to junior high school: A longitudinal study of self-esteem, psychological symptomatology, school life, and social support. *Child Development, 58,* 1235–1243. http://dx.doi.org/10.2307/1130617

Hitz, R., & Wright, D. (1988). Kindergarten issues: A practitioners' survey. *Principal, 67,* 28–30.

Huston, A. C., & Bentley, A. C. (2010). Human development in societal context. *Annual Review of Psychology, 61,* 411–437.

Huttenlocher, J., Vasilyeva, M., Waterfall, H. R., Vevea, J. L., & Hedges, L. V. (2007). The varieties of speech to young children. *Developmental Psychology, 43,* 1062–1083. http://dx.doi.org/10.1037/0012-1649.43.5.1062

Ing, M., Webb, N. M., Franke, M. L., Turrou, A. C., Wong, J., Shin, N., & Fernandez, C. H. (2014). *How student participation mediates the relationship between teacher practices and student achievement*. Paper presented at the annual meeting of the American Educational Research, Philadelphia, PA.

Jackson, K., & Wilson, J. (2012). Supporting African American students' learning of mathematics: A problem of practice. *Urban Education, 47,* 354–398. http://dx.doi.org/10.1177/0042085911429083

Jussim, L. (1989). Teacher expectations: Self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and Social Psychology, 57,* 469–480. http://dx.doi.org/10.1037/0022-3514.57.3.469

Jussim, L., Eccles, J., & Madon, S. (1996). Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling prophecy. *Advances in Experimental Social Psychology, 28,* 281–388. http://dx.doi.org/10.1016/S0065-2601(08)60240-3

Klein, A., Starkey, P., & Wakeley, A. (2000). *Child Math Assessment: Preschool Battery (CMA)*. Berkeley, CA: University of California.

Kozol, J. (1991). *Savage inequalities: Children in America's schools*. New York, NY: Crown Publishers.

Kuklinski, M. R., & Weinstein, R. S. (2001). Classroom and developmental differences in a path model of teacher expectancy effects. *Child Development, 72,* 1554–1578. http://dx.doi.org/10.1111/1467-8624.00365

Ladson-Billings, G. (1997). It doesn't add up: African American students' mathematics achievement. *Journal for Research in Mathematics Education, 28,* 697–708. http://dx.doi.org/10.2307/749638

Lee, V. E., & Burkam, D. T. (2002). *Inequality at the starting gate: Social background differences in achievement as children enter school*. Washington, DC: Economic Policy Institute.

Loeb, S., & Bassok, D. (2008). Early childhood and the achievement gap. In H. F. Ladd & E. B. Fiske (Eds.), *Handbook of research in education policy finance and policy* (pp. 517–534). New York, NY: Routledge.

Lopez, N. (2002). *Hopeful girls, troubled boys: Race and gender disparity in urban education*. New York, NY: Routledge.

Lubienski, S. T. (2002). A closer look at Black-White mathematics gaps: Intersections of race and SES in NAEP achievement and instructional practices data. *Journal of Negro Education, 71,* 269–287. http://dx.doi.org/10.2307/3211180

Lubienski, S. T. (2006). Examining instruction, achievement, and equity with NAEP mathematics data. *Education Policy Analysis Archives, 14,* 1–33.

Madon, S., Jussim, L., Keiper, S., Eccles, J., Smith, A., & Palumbo, P. (1998). The accuracy and power of sex, social class, and ethnic stereotypes: A naturalistic study in person perception. *Personality and Social Psychology Bulletin, 24,* 1304–1318. http://dx.doi.org/10.1177/01461672982412005

McKown, C., & Weinstein, R. S. (2008). Teacher expectations, classroom context, and the achievement gap. *Journal of School Psychology, 46,* 235–261. http://dx.doi.org/10.1016/j.jsp.2007.05.001

Midgley, C., Feldlaufer, H., & Eccles, J. S. (1989). Change in teacher efficacy and student self-and task-related beliefs in mathematics during the transition to junior high school. *Journal of Educational Psychology, 81,* 247–258. http://dx.doi.org/10.1037/0022-0663.81.2.247

Muthén, L. K., & Muthén, B. O. (2013). Mplus version 7.1 [Computer software]. Los Angeles, CA: Author.

National Center for Education Statistics. (2013). *The nation's report card: Trends in academic progress 2012*. Washington, DC: Institute of Education Sciences, U.S. Department of Education.

National Council of Teachers of Mathematics. (1991). *1991–1992 handbook: NCTM goals, leaders, and positions*. Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education.

National Research Council. (2009). *Mathematics in early childhood: Learning paths toward excellence and equity*. Washington, DC: National Academy Press.

NICHD Early Child Care Research Network. (1996). Characteristics of infant child care: Factors contributing to positive caregiving. *Early Childhood Research Quarterly, 11,* 269–306.

Oakes, J. (1990). *Multiplying inequalities: The effects of race, social class, and ability grouping on opportunities to learn math and science*. Santa Monica, CA: RAND Corporation.

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62,* 307–332. http://dx.doi.org/10.3102/00346543062003307

Peisner-Feinberg, E. S., & Burchinal, M. R. (1997). Relations between preschool children's child-care experiences and concurrent development: The Cost, Quality, and Outcomes Study. *Merrill-Palmer Quarterly, 43,* 451–477.

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

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods, 15,* 209–233. http://dx.doi.org/10.1037/a0020141

Proctor, N. (1984). Towards a partnership with schools. *Journal of Education for Teaching, 10,* 219–232. http://dx.doi.org/10.1080/0260747840100303

Raudenbush, S. W. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy induction: A synthesis of findings from 18 experiments. *Journal of Educational Psychology, 76,* 85–97. http://dx.doi.org/10.1037/0022-0663.76.1.85

Reardon, S. F., & Galindo, C. (2006). *Patterns of Hispanic students' math and English literacy test scores in the early elementary grades*. National Task Force on Early Childhood Education for Hispanics.

Reardon, S. F., & Robinson, J. P. (2008). Patterns and trends in racial/ethnic and socioeconomic academic achievement gaps. In H. F. Ladd & E. B. Fiske (Eds.), *Handbook of research in education finance and policy* (pp. 497–516). New York, NY: Routledge.

Reardon, S. F., Robinson-Cimpian, J. P., & Weathers, E. S. (2015). Patterns and trends in racial/ethnic and socioeconomic academic achievement gaps. In H. Ladd & M. Goertz (Eds.), *Handbook of research in education finance and policy* (2nd ed., pp. 491–509). Mahwah, NJ: Erlbaum.

Retelsdorf, J., Schwartz, K., & Asbrock, F. (2015). "Michael can't read!" Teachers' gender stereotypes and boys' reading self-concept. *Journal of Educational Psychology, 107,* 186–194. http://dx.doi.org/10.1037/a0037107

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review, 3,* 16–20. http://dx.doi.org/10.1007/BF02322211

Rubie-Davies, C. M. (2007). Classroom interactions: Exploring the practices of high- and low-expectation teachers. *British Journal of Educational Psychology, 77,* 289–306. http://dx.doi.org/10.1348/000709906X101601

Rubie-Davies, C. M., Peterson, E. R., Sibley, C. G., & Rosenthal, R. (2015). A teacher expectation intervention: Modeling the practices of high expectation teachers. *Contemporary Educational Psychology, 40,* 72–85. http://dx.doi.org/10.1016/j.cedpsych.2014.03.003

Rubie-Davies, C. M., Weinstein, R. S., Huang, F. L., Gregory, A., Cowan, P. A., & Cowan, C. P. (2014). Successive teacher expectation effects across the early school years. *Journal of Applied Developmental Psychology, 35,* 181–191. http://dx.doi.org/10.1016/j.appdev.2014.03.006

Rumberger, R., & Palardy, G. (2005). Does segregation still matter? The impact of student composition on academic achievement in high school. *Teachers College Record, 107,* 1999–2045.

Ruzek, E. A., Hafen, C. A., Allen, J. P., Gregory, A., Mikami, A. Y., & Pianta, R. C. (2016). How teacher emotional support motivates students: The mediating roles of perceived peer relatedness, autonomy support, and competence. *Learning and Instruction, 42,* 95–103. http://dx.doi.org/10.1016/j.learninstruc.2016.01.004

Saft, E. W., & Pianta, R. C. (2001). Teachers' perceptions of their relationships with students: Effects of child age, gender, and ethnicity of teachers and children. *School Psychology Quarterly, 16,* 125–141. http://dx.doi.org/10.1521/scpq.16.2.125.18698

Sarama, J., & Clements, D. H. (2013). Lessons learned in the implementation of the TRIAD scale-up model: Teaching early mathematics with trajectories and technologies. In T. G. Halle, A. J. Metz, & I. Martinez-Beck (Eds.), *Applying implementation science in early childhood programs and systems* (pp. 173–191). Baltimore, MD: Brookes.

Sarama, J., Clements, D. H., Starkey, P., Klein, A., & Wakeley, A. (2008). Scaling up the implementation of a pre-kindergarten mathematics curriculumml: Teaching for understanding with trajectories and technologies. *Journal of Research on Educational Effectiveness, 1,* 89–119. http://dx.doi.org/10.1080/19345740801941332

Sarama, J., Clements, D. H., Wolfe, C. B., & Spitler, M. E. (2012). Longitudinal evaluation of a scale-up model for teaching mathematics with trajectories and technologies. *Journal of Research on Educational Effectiveness, 5,* 105–135. http://dx.doi.org/10.1080/19345747.2011.627980

Sarama, J., & DiBiase, A-M. (2004). The professional development challenge in preschool mathematics. In D. H. Clements, J. Sarama, & A.-M. DiBiase (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education* (pp. 415–446). Mahwah, NJ: Erlbaum.

Schoen, H. L., Cebulla, K. J., Finn, K. F., & Fi, C. (2003). Teacher variables that relate to student achievement when using a standards-based curriculum. *Journal for Research in Mathematics Education, 34,* 228–259. http://dx.doi.org/10.2307/30034779

Selig, J. P., & Preacher, K. J. (2008). Monte Carlo method for assessing mediation: An interactive tool for creating confidence intervals for indirect effects [Computer software]. Retrieved from http://www.quantpsy.org

Seo, K., & Ginsburg, H. P. (2004). What is developmentally appropriate in early childhood mathematics education? Lessons from new research. In D. H. Clements & J. Sarama (Eds.), *Engaging young children in mathematics: Standards for early mathematics education* (pp. 91–104). Mahwah, NJ: Erlbaum.

Silver, E. A., & Stein, M. K. (1996). The Quasar Project: The "revolution of the possible" in mathematics instructional reform in urban middle schools. *Urban Education, 30,* 476–521. http://dx.doi.org/10.1177/0042085996030004006

Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroomml: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology, 85,* 571–581. http://dx.doi.org/10.1037/0022-0663.85.4.571

Sonnenschein, S., & Galindo, C. (2015). Race/ethnicity and early mathematics skills: Relations between home, classroom, and mathematics achievement. *The Journal of Educational Research, 108,* 261–277. http://dx.doi.org/10.1080/00220671.2014.880394

Steele, C. M. (1997). A threat in the air. How stereotypes shape intellectual identity and performance. *American Psychologist, 52,* 613–629. http://dx.doi.org/10.1037/0003-066X.52.6.613

Stinson, D. W. (2006). African American male adolescents, schooling (and mathematics): Deficiency, rejection, and achievement. *Review of Educational Research, 76,* 477–506. http://dx.doi.org/10.3102/00346543076004477

Stipek, D. (1998). Motivation and instruction. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 85–113). New York, NY: Macmillan.

Stipek, D., & Byler, P. (2004). The early childhood classroom observation measure. *Early Childhood Research Quarterly, 19,* 375–397. http://dx.doi.org/10.1016/j.ecresq.2004.07.007

Stipek, D. J., Givvin, K. B., Salmon, J. M., & MacGyvers, V. L. (2001). Teachers' beliefs and practices related to mathematics instruction. *Teaching and Teacher Education, 17,* 213–226. http://dx.doi.org/10.1016/S0742-051X(00)00052-4

Stipek, D. J., & Ryan, R. H. (1997). Economically disadvantaged preschoolers: Ready to learn but further to go. *Developmental Psychology, 33,* 711–723. http://dx.doi.org/10.1037/0012-1649.33.4.711

Strutchens, M. E., & Silver, E. A. (2000). NAEP findings regarding race/ethnicity: Students' performance, school experiences, and attitudes and beliefs. In E. A. Silver, & P. A. Kenny (Eds.), *Results from the seventh mathematics assessment of the National Assessment of Educational Progress* (pp. 45–72). Reston, VA: National Council of Teachers of Mathematics.

Thompson, A. G. (1984). The relationship of teachers' conceptions of mathematics and mathematics teaching to instructional practice. *Educational Studies in Mathematics, 15,* 105–127. http://dx.doi.org/10.1007/BF00305892

Walshaw, M., & Anthony, G. (2008). The teacher's role in classroom discourse: A review of recent research into mathematics classrooms. *Review of Educational Research, 78,* 516–551. http://dx.doi.org/10.3102/0034654308320292

Weinstein, R. S., Soulé, C. R., Collins, F., Cone, J., Mehlhorn, M., & Simontacchi, K. (1991). Expectations and high school change: Teacher-researcher collaboration to prevent school failure. *American Journal of Community Psychology, 19,* 333–363.

Wenglinsky, H. (2004). Closing the racial achievement gap: The role of reforming instructional practices. *Education Policy Analysis Archives, 12*(64), 1–22.

West, C. K., & Anderson, T. H. (1976). The question of preponderant causation in teacher expectancy research. *Review of Educational Research, 46,* 613–630. http://dx.doi.org/10.3102/00346543046004613

Wigfield, A., Eccles, J. S., Mac Iver, D., Reuman, D. A., & Midgley, C. (1991). Transitions during early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology, 27,* 552–565. http://dx.doi.org/10.1037/0012-1649.27.4.552

Received February 8, 2016
Revision received October 12, 2016
Accepted October 19, 2016 ■