Implementation of Elementary Mathematics Materials:
A Longitudinal Analysis of School-Aggregated Fourth-Grade Achievement Outcomes

Jessica M. Young, Kristen E. Reed, Deborah B. Spencer, Louisa Anastasopoulos, June Mark, Katherine Schwinden
Education Development Center, Inc., Waltham, Massachusetts

This research was funded by the National Science Foundation, award number 0918109. We would like to acknowledge the following colleagues who assisted with conducting the study and critiquing the manuscript: Laura O’Dwyer, Julie Zeringue, Deborah Rosenfeld, Mari Halladay, Jess Gropen, and Kathryn Chiappinelli.

Correspondence concerning this article should be addressed to Jessica M. Young, Education Development Center, Inc., 43 Foundry Ave., Waltham, MA 02453.
Email: jyoung@edc.org
Abstract

How do district-led elementary mathematics improvement efforts relate to changes in fourth-grade students’ achievement as measured by the state test? We describe a multilevel mixed methods project studying 10 districts’ implementation of two elementary mathematics materials. We employed a multilevel piece-wise growth curve analysis using six years of school aggregated assessment data—three years prior to the adoption of the new materials and three following adoption. Results suggest a potential effect of well-supported interventions on student outcomes, where schools that received higher levels of support from their districts on average had students who grew significantly more in fourth-grade mathematics achievement between time points compared with schools that received less support even after controlling for students’ low-income status. The results may assist administrators, as well as policymakers, in planning an implementation of new mathematics materials. This project contributes to education research by using a longitudinal quasi-experimental analysis to study curriculum implementation.
Implementation of Elementary Mathematics Materials:
A Longitudinal Analysis of School Aggregated Fourth Grade Achievement

Over the past decade, many school districts across the country have centered their mathematics program improvement efforts—especially at the elementary level—on the adoption and use of instructional materials well-aligned to state standards (Mark, Spencer, Zeringue, & Schwinden, 2010). These district-level improvement efforts are a response to increased accountability for student outcomes in mathematics, and are intended to ensure that all students experience high-quality mathematics content and instruction. If supported well by districts and schools, instructional materials can be a driver for mathematics program improvement, particularly when the materials are used consistently across schools and grade levels, and well-aligned with the standards and tests for which the district is held accountable (Briars & Resnick, 2000; Cohen & Hill, 2001; Porter, Smithson, Blank, & Zeider, 2007; Roach, Niebling & Kurz, 2008; Schmidt, Houang, & Cogan, 2002). Further supporting this movement is the emergence of well-designed instructional materials—including several programs developed with funding from the National Science Foundation—that build on research about student learning of mathematics, respond to national mathematics standards detailing increased expectations for student learning, and strive to support teacher learning as well as student learning. Programs such as Everyday Mathematics, 3rd edition, ©2007 and Investigations in Number, Data and Space, 2nd edition, ©2008 are now widely used across the country.

Researchers have long been in agreement that instructional materials matter to student outcomes (Begle, 1973; Begle, 1979; Driscoll, 1980; Glidden, 1991; McKnight et al., 1987; Porter, 1989; Robitaille & Travers, 1992; Schmidt, McKnight, & Raizen, 1997; Tyson, 1997;
Usiskin, 1985). Four decades ago, the School Mathematics Study Group provided this summation of the National Longitudinal Study of Mathematical Abilities: “If a mathematical topic is in the text, then students do learn it. If the topic is not in the text, then, on average, students do not learn it” (Begle, 1973, p. 209). The Third International Mathematics and Science Study (TIMSS) showed that this pattern endures; the performance of U.S. eighth graders was notably different on different topic areas, and those differences could be linked to differences in emphasis in instructional materials. The correlation between topics that were covered in the textbook and what teachers actually taught was a stunning .95. One of the most important findings of TIMSS is “that the curriculum itself—what is taught—makes a huge difference” (Schmidt et al., 2002, p. 12, authors’ italics).

Instructional materials can play a defining role in mathematics classrooms, affecting both what and how teachers teach (National Research Council, 2002; Reys, Reys, Lapan, & Holliday, 2003; Van Zoest & Bohl, 2002). Ball and Cohen (1996) explain this influence: “Unlike frameworks, objectives, assessments, and other mechanisms that seek to guide curriculum, instructional materials are concrete and daily. They are the stuff of lessons and units, of what teachers and students do…. Not only are curriculum materials well-positioned to influence individual teachers’ work but, unlike many other innovations, textbooks are already ‘scaled up’ and part of the routine of schools. They have ‘reach’ in the system” (p. 6). When ignored, instructional materials can become a stumbling block. Schmidt et al. (2002) cautions against efforts to improve instruction that are isolated from efforts to improve instructional materials: “If we pretend that the textbook doesn’t exist—and conduct professional development in ways that assume teachers will implement an entirely different approach to content than the texts take—
believe me, the textbook will win” (p. 18). To successfully strengthen mathematics instruction, school districts must consider how materials will play a part.

The promise of such materials is great, but their impact is mediated by how well they are used by teachers in their classrooms (Stein, Remillard, & Smith, 2007). Materials are an expression of the intended curriculum (McKnight et al., 1987), but the curriculum as enacted in the classroom is “actually jointly constructed by teachers, students, and materials in a particular context” (Ball & Cohen, 1996, p. 7). Teachers adapt the materials to suit their particular students’ needs and their local context—and those adaptations also reflect their own understanding of and beliefs about the subject matter and about how children learn that content (Ball, 2003; Ball & Cohen, 1996; Boyd, Banilower, Pasley, & Weiss, 2003; Cohen & Ball, 1999; Davenport, 2000; Kilpatrick, 2003; National Research Council, 2001; Stein, Smith, & Silver, 1999). Instructional materials do not stand alone: they must be carefully aligned with other elements of practice and policy at the district and school levels to provide teachers with the coherent instructional guidance (Spillane, 1998) needed to enact the materials’ vision for classroom practice (Boyd et al., 2003; Briars & Resnick, 2000; Char, 2004; Lord et al., 2000; Schoenfeld, 2002; Stein & Coburn, 2008; Swafford & Langrall, 2008).

In this paper, we describe a four-year, NSF-funded study that is investigating the effects of a district-level strategy for scaling up instructional improvement that is centered on the consistent use of elementary mathematics materials. Our project, Coherent Implementation of Mathematics Instructional Materials: A Study in the Variations and Effects of District Supports for Implementation (ESI-0918109) employs a mixed-methods, multilevel longitudinal design to study the implementation of Everyday Mathematics (EDM) and Investigations in Number, Data, and Space (INV) in 12 school districts across five states. Our study explores the relation among
the district-level support provided for implementation, the school-level support provided for implementation, the level of use of the materials at the school level, and the effect on student outcomes (as shown in Figure 1).

Our overall hypothesis, building on the literature in the field and the results of our prior work (Mark et al., 2010), is that school districts that provide higher levels of support for curriculum implementation—guidance on how and when materials are used, accountability for use, learning opportunities for teachers and principals, sufficient resources, and consistent messages across levels of the system—will see stronger student outcomes. We expect that district-level support for implementation, when paired with support for implementation at the school level, will result in greater use of the instructional materials at the classroom level, ultimately leading to increased student outcomes. Conversely, we expect that a low level of support for implementation from the district and the school would be correlated with a low level of classroom use, and would be less likely to have a positive effect on student outcomes. These central questions have focused the research:

1. What are the dimensions of district-level support for implementation of instructional materials?
2. What are the dimensions of school-level support for implementation of instructional materials?
3. Does district-level support relate to changes in school-level student mathematics scores, at a single point in time and over time?
4. Does school-level support for implementation and/or school-level use mediate the relation between district-level support and student outcomes?

In this paper, we investigate and operationalize the key dimensions of support that constitute a
coherent implementation strategy from a district perspective and examine the connection between district-level support and student outcomes as measured on state assessments (see Figure 1). We first report on our method for deriving and confirming the dimensions of district-level support for implementation. Next, we examine whether the implementation of new mathematics materials impact school-aggregated fourth-grade state mathematics test scores over six years—three years before implementation and three years following implementation. Finally, we examine whether districts’ level of support promotes growth in student achievement during the first three years of implementation. Student outcomes are measured by school-level progress toward mathematical proficiency on annual state tests, using a multilevel piece-wise growth curve analysis. Results should be of major interest to school administrators and policymakers across the United States as they plan, fund, and implement efforts to use new instructional materials to advance mathematics program improvement—shedding light on factors affecting the use of such materials at large scale. We hope our results will be particularly useful as districts consider investing in improvement strategies to address the Common Core State Standards for Mathematics (National Governors Association & Council of Chief State School Officers, 2010) and determine the role instructional materials will play in those efforts. We conducted two studies in order to examine these specific research questions:

1. What are the dimensions of district-level support for implementation of instructional materials?
2. Does the implementation of new mathematics materials increase students’ mathematics achievement over time? Are there differences among schools in the growth of the pre-implementation mathematics achievement, mathematics
achievement in the year prior to implementation, and post-implementation mathematics achievement?

3. Does the percentage of low-income students predict differences among schools in the growth of their mathematics achievement over time?

4. After controlling for the percentage of low-income students in the school, does the level of support that districts provide to schools predict differences among schools in the growth of their post-implementation mathematics achievement scores?

Study 1

Study 1 addressed the first research question: What are the dimensions of district-level support for the implementation of instructional materials? We undertook this qualitative interview study because of the importance of the construct of district-level support to the overall research, and to reduce validity threats and increase the credibility of our findings.

Ensuring the adopted materials are used and used well is a significant challenge. Our previous research has shown that school districts vary in the degree to which they provide and align the supports necessary for such a coherent implementation (Mark et al., 2010). Based on prior work (Mark et al., 2010) and a review of the relevant literature, we initially identified seven dimensions of support that districts provide to schools: (1) consistent use of instructional materials across all schools that were well-aligned to state standards; (2) guidance on how materials are used and when; (3) accountability for use of the materials; (4) professional learning opportunities for teachers that support curriculum use; (5) professional learning opportunities for principals that support curriculum use; (6) sufficient resources to teach the intended program; (7)
minimal distracters or competing initiatives. To validate and expand these initial dimensions we interviewed a sample of district-level curriculum leaders.

Method

Sample. The sample consisted of 23 district leaders from 21 districts in 11 different states, interviewed between May and November 2010. The 21 districts ranged in size from 2,000 to 56,000 students, with the majority enrolling approximately 10,000 students in fewer than 20 elementary schools. The percentage of low-income students in the districts ranged from a low of 7% to a high of 77%; the average percentage of low-income students across the sample was 45%. Nine districts were urban; nine were suburban—within commuting distance of larger cities; and two were rural. All had implemented either INV or EDM within the previous 10 years. The districts of these curriculum leaders met many of the criteria for full-study participation (described in Study 2), but were not able to participate in our final sample (typically because the year of mathematics materials adoption in the district was not 2008 or 2009, as Study 2 [the full study] requires).

Research Design. We interviewed district leaders over the phone for approximately an hour about their experiences supporting implementation in their districts. The interview protocol began with an open-ended question asking participants to describe their districts’ mathematics materials implementation. This was to ensure that we did not impose our framework on their responses, in an effort to limit potential bias in their account and our interpretation (Maxwell, 1996). We then followed a semi-structured interview protocol designed to engage district leaders with specific questions related to the dimensions of district-level support that we had identified and to see how these dimensions played a role in their districts’ implementation. These questions were intended to examine whether the dimensions of coherent implementation that we had
identified *a priori* were valid in their experience. Miles and Huberman (1994) refer to this as “truth value.” Finally, we probed for additional dimensions of support for the implementation that we missed or had not discussed.

The qualitative data were analyzed with regard to the proposed dimensions, testing the validity of the construct of district-level support by looking for discrepant cases. Each interview was transcribed and summarized in an analytic memo. The memo identified salient features of the district’s implementation with careful attention to features that were not included in our initial set of dimensions. Each memo included a chart that detailed whether and how our identified dimensions played a role in that district’s implementation. Where necessary, additional dimensions were added or the definitions of dimensions were expanded. We also coded each interview using qualitative analysis software. The first set of interviews (10) was coded by sorting the data into our dimensions in order to understand what the range of data looked like in any given dimension. The remaining transcripts were coded more holistically, looking for themes and trends, as well as adding to the range of data in each dimension.

**Results**

This analysis confirmed that the seven dimensions of district-level support we had initially identified were valid. The findings of this study allowed us to clarify some of the key aspects of the dimensions and provided us with an in-depth understanding of how districts may vary in their support for implementation. Specifically, we added more depth to the dimension of “accountability for use of the materials,” noting that districts may have a system for monitoring classroom instruction and providing follow-through on identified problems (or broadcasting successes). We renamed the first dimension from “consistent use of instructional materials across all schools that were well-aligned to state standards” to “alignment to standards.”
Administrators at the district-level did have responsibility for implementing materials aligned to the state standards, but while district leaders could set the expectation for consistent use, actual use occurs at the school level and varies by school (and sometimes by classroom). (This construct of “use of the materials” will be analyzed as a school-level variable in future papers as it was measured in the last year of the study.) We also expanded and renamed our last dimension of “minimal distracters or competing initiatives” to “consistent message and vision” so that we could include alignment between district departments and staff, particularly those who are “in the line,” supervising principals and math support personnel. Study 1 further revealed that problems arose in districts when a district leader who supervised principals (such as an assistant superintendent) was not supportive of the mathematics adoption (or, in extreme cases, spoke out against it), contradicting the math department’s message that “these materials will help our district be more successful and everyone should use them.” The revised dimensions of district-level support for elementary mathematics materials implementation developed in this study include:

1. **Alignment to standards.** During the selection and implementation, district leaders and teachers evaluate and continue to monitor that the materials are aligned with their state standards, and adjust as needed.

2. **Guidance on how materials are used and when.** Pacing guides or curriculum maps suggest the order and timing of use of units or lessons, and may also advise on the key activities in the units taught, and the use of assessments or homework. These tools encourage greater use and greater fidelity to the materials (Choppin, 2006; Goldsmith, Mark, & Kantrov, 2000; Mark et al., 2010).

3. **Accountability for use of the materials and clear expectations for use.** The district
communicates to teachers and principals that the expectation is that all teachers in all elementary schools will use *Everyday Mathematics* or *Investigations* as their primary instructional materials.

4. **Professional learning opportunities for teachers.** Professional development that familiarizes teachers with the content and instructional approaches of mathematics materials—including orientations, unit walkthroughs, grade-level meetings, and classroom coaching—is essential for effective implementation (Bay, Reys, & Reys, 1999; Goldsmith et al., 2000; Penuel, Fishman, Yamaguchi, & Gallagher, 2007; Stein & Kim, 2009). Quality curriculum-related professional development also strengthens teachers’ instructional practice and deepens their knowledge of mathematics content, pedagogy, and assessment (Davenport, 2000; Empson & Junk, 2004; Phillips, Lappan, Grant, & Arbaugh, 2008; Stein et al., 2007).

5. **Professional learning opportunities for principals.** To support implementation in their schools, principals must learn about the materials—about both the underlying approaches to mathematics and student learning (Coburn, 2005; Nelson, 1999; Spillane, Reiser, & Reimer, 2002) and the supports teachers need to use them well.

6. **Sufficient resources.** Teachers must have the needed materials to use the program well (Center for Applied Research and Educational Improvement, 2001). Resources also include instructional time (Keiser & Lambdin, 1996): if materials specify that 60 minutes of daily mathematics time is needed, schedules must reflect that or teachers are not able to use the program as designed.

7. **Consistent message and vision.** There must be a consistent message and shared vision through all levels of the school and district about the expectations for implementation.
Teachers who receive the same message about how to use the materials from their principal as they do from the district math director and superintendent are more likely to invest the time and energy required to learn to use the materials well. This consistent vision also allows resources and supports to be aligned in service of implementation.

Discussion

Past research has shown that instructional materials matter in terms of what is taught in the classroom and what students learn, but less is known about how structures, policies, and decisions at the district level support teachers’ use of the materials and ultimately student achievement. This qualitative study begins to answer the question of what are the district-level supports for implementation by providing a framework with seven dimensions. In Study 2, we applied these seven dimensions to the districts in the larger study, analyzing the relation between the level of district support and growth in school-aggregated fourth-grade student achievement.

Study 2

Study 2 examined two research questions: (1) Does the implementation of new mathematics materials increase school-aggregated fourth-grade mathematics achievement over time? (2) Does the level of support that districts provide to schools predict differences among schools in the growth of their post-implementation aggregated achievement scores? We studied districts that implemented new mathematics materials in all of their elementary schools in either the 2008–2009 school year (cohort 1) or the 2009–2010 school year (cohort 2). Three years after the initial year of implementation (2010–11 for cohort 1 and 2011–12 for cohort 2), teachers and administrators were surveyed in each school about their use of the materials and support for implementation, and a purposeful subsample of teachers was interviewed about their experience with the materials. Additional information about the supports for implementation at the district
level was collected through interviews with district leaders, review of district documents, and site visits. Average fourth-grade student achievement data for each school was compiled starting three years prior to implementation of the mathematics materials through three years of implementation.

Methods

Sample. Longitudinal school achievement data was obtained from publicly available state datasets for 152 participating schools representing 12 districts in five states in the United States. One state assessment agency could only provide scaled student achievement scores, and after examining the data, it was determined that the data violated the even-interval-level assumption; because the even-interval assumption is fundamental to calculating means and standard deviations, we could not include the data. Therefore, the longitudinal data of 131 schools representing 10 districts in four states were included in this analysis. Additional survey data were collected from approximately 2,000 teachers and 131 principals, and 115 interviews were conducted with teachers, principals and district leaders during the third year of implementation. At least one site visit was conducted in each district.

The districts participating in this study met several selection criteria. First, each district in the sample implemented one of two elementary mathematics curricula—*Everyday Mathematics* or *Investigations in Number, Data, and Space*—in all of their elementary schools. The sample was limited to districts using one of these two programs in order to reduce the variability due to program choice, and to enable generalizability beyond one particular set of instructional materials. These two programs were chosen because they are high-quality, standards-based instructional materials programs that are widely used in the United States. In addition, these two programs share with the *Common Core* a goal of bringing greater focus and coherence to K–12
mathematics, an emphasis on mathematical practices, and attention to developing conceptual understanding and important skills. Both programs were developed with NSF support, which allowed their authors to attend to student learning without undue influence of the market pressures that drive materials development by commercial publishers. Each program was designed by interdisciplinary teams of mathematics educators, mathematicians, and teachers; aligned to state and national standards; and extensively field tested. This study was not designed to compare these two programs but, rather, to consider them variations of a similar strategy.

Both programs have been found to have a positive effect on student learning. *Everyday Mathematics* has the largest research base (see, for example, NRC, 2004; Riordan & Noyce, 2001; Sconiers, McBride, Issacs, Kelso, & Higgins, 2003; SRA/McGraw Hill, 2003; Waite, 2000), and it has been given a positive rating by the What Works Clearinghouse. *Investigations* has also been given a positive rating by the What Works Clearinghouse. The ARC Center Tri-State Student Achievement Study (Sconiers et al., 2003)—which included over 100,000 students—found that students using *Everyday Mathematics, Investigations, and Math Trailblazers* consistently outperformed students using comparison curricula across all considered tests, grade levels, and strands.

District size was limited to a total student population of 3,000–35,000. This was established to ensure that districts were not so small that they did not need to establish a strategy for implementation across multiple elementary schools, but also not so large that the district-level efforts to support implementation would be difficult to observe, as very large districts might require an additional layer of bureaucracy in order to manage curriculum and instruction. Finally, participating districts had to be in states that had stable state standards and a stable state assessment system from 2005 to 2012. This was particularly important as the student
achievement data were gathered from state mathematics test results for fourth graders, aggregated to the school level.

The 131 elementary schools from these districts range in their demographic variables including urbanicity, poverty levels, and student race/ethnicity. Schools are in rural and suburban areas, large and midsize cities, and remote towns. As reported by the National Center for Educational Statistics (NCES), in the 2011–12 school year, on average 55% of the students in the sample received free/reduced lunch (SD 27%), and 65% of the schools are Title I eligible while 52% are Schoolwide Title I (Keaton, 2012). The average school size is 504 students with a standard deviation of 135. The elementary schools educated over 66,000 students in the 2011–12 school year. The racial/ethnic breakdown of the overall sample (all students in the elementary schools, not just grade 4) is <1% American Indian/Alaskan Native; 8% Asian; 41% Hispanic; 10% Black; 33% White; <1% Hawaiian Native/Pacific Islander; 7% two or more races. Table 1 illustrates the demographics for each of the 10 districts.

**Instruments and Measures.** Below we describe the longitudinal analysis of the student achievement measures, school and district characteristics, and district-level support instruments and measures. Additional data about implementation at the school level were collected—through surveys and interviews with teachers and principals— the results of which will be reported in a subsequent article.

**Longitudinal Mathematics Achievement Scores.** Six years of fourth-grade-student achievement data in mathematics was collected from state assessment agencies for each of our 131 schools and standardized to compare across state tests. Three years of post-implementation achievement scores was collected, culminating in 2011 for cohort 1 and 2012 for cohort 2 because we expected the effects of implementation on student learning to be more visible after
three years. This period of time also allowed teachers to become adept in the use of the materials, and students to experience the program in multiple grades. State assessment agencies provided publicly available fourth-grade students’ mathematics raw scores on the state test from 2006 to 2011 for schools in cohort 1 and from 2007 to 2012 for cohort 2 schools. The data did not contain any personally identifying information, aside from the students’ school and grade level. In their raw form, these data are not directly comparable, as each state (at the time of this report) designs its own test and sets its own level of proficiency on that test. Therefore, we used a z-score methodology to standardize the school-level achievement scores centering on each state’s minimum proficiency and dividing by the standard deviation of all schools in that state. This creates a score with a mean of zero and a standard deviation of 1, and enables direct comparison of the assessment data across all schools in the sample. Each state’s raw scores are linear and set on an equal interval, which is a key assumption in the z-score calculation. A z-score for each school was calculated using the following formulas (j = school; k = district):

\[
Z_j = \frac{(X_j \text{ average raw score of school } j) - X_{jk} \text{ (min raw score proficiency across all schools in state)}}{S_{jk} \text{ (SD across all schools in state)}}
\]

While z-scores allow us to use a common scale for comparing scores across states and time points, it is important to note that since the z-scores are calculated using the minimum proficiency scores for each state, z-scores for a state with lower standards for “proficiency” will be higher than z-scores for a state with more stringent proficiency requirements. For example, the Massachusetts schools in our sample had the lowest mean z-scores; the participating Pennsylvania schools had the highest, and the California and Texas schools fell somewhere between the two (see Figure 2). This disparity is in part a function of the different standards for proficiency in each state. A comparison of grade-4 mathematics standards across states conducted as part of the National Assessment of Educational Progress (NAEP; Bandeira de
Mello, 2011), which relates converted state minimum proficiency scores to NAEP scale equivalents, shows this disparity (see Figure 3). Of the states in our study, Massachusetts is the only one with a proficiency level higher than the NAEP standard. The proficiency levels for California, Pennsylvania, and Texas fall closer to what NAEP would classify as a “Basic” level of mathematical understanding. Given the disparity in proficiency levels across the states in our sample, we caution that it is not useful or valid to compare initial status of z-scores between states; rather the focus of this paper is on the change in z-scores over time.

**School Characteristics.** To have a directly comparable measure of low income for each of the schools in the sample, we compiled data from the 2010–11 NCES Common Core of Data (Keaton, 2012). This coincided with the third year of implementation for schools in cohort 1. The data for the third year of implementation for cohort 2 (2011–12) is not yet available from NCES. To establish that the percent of low-income students was fairly stable across time, we compared the 2008–09 data with the 2010–11 data using a paired samples t-test. The results indicated no significant difference in the percent of free and reduced lunch between the 2009 data and the 2011 data. See the district demographic table (Table 1) for district-level characteristics of each district, such as race/ethnicity, number of elementary schools, number of students in the district, and number of elementary school students.

**District-Level Support.** During the third year of implementation in each of the 10 sample districts, we collected qualitative data to be used as a measure of the supports provided at the district level for curriculum implementation. We interviewed two or three district-level administrators (typically, mathematics administrators and supervisors of principals, often a superintendent or an assistant superintendent), conducted a site visit, and collected district-level documents (primarily pacing guides and other artifacts of curriculum planning and support). For
the interview sample, we aimed to include the math director and the supervisor of principals (typically, an assistant superintendent or a superintendent). In two districts, we were only able to interview one district administrator (the administrator responsible for mathematics instruction) and included interviews with principals (three in each district) to supplement the missing data. The interviews followed a similar protocol to Study 1. District leaders’ were first asked an open-ended question about their experiences supporting curriculum implementation over the course of the three-year implementation period, followed by a semi-structured protocol that targeted the specific dimensions of support identified in Study 1. These included questions about the provision of basic supports like the materials themselves and instructional time needed to teach each program; their efforts to build the capacity of teachers and administrators to use the materials well; and their efforts to create accountability for and monitor use of the materials.

All 22 district-level interviews were audio recorded and transcribed. A researcher reviewed each transcript, looking for evidence of the dimensions of district-level support described in our theoretical framework, and summarized the interview in a table according to the dimensions. Because our framework for district-level support addresses a particular set of pre-defined categories, we did not code the data using grounded-theory. Rather, we analyzed the transcripts looking for patterns, themes, and outliers that explained the presence or absence of particular elements of support. The summarized interviews for each district were then aggregated in a district summary memo, along with notes from site visits and the analysis of district documents. These data sources were compared for confirming and disconfirming evidence. When disconfirming evidence was found, researchers went back to the transcripts or other original documents to try to resolve the difference, or when there continued to be a conflict, noted the discrepancy. Two other researchers read and critiqued the district summary memo; the
critique was then used to revise and strengthen the memos. Next, the data were compiled into a matrix to compare the dimensions of support across all 10 districts. Using this 10-district matrix, the level of support was coded across all of the support dimensions—generally testing for the presence or absence of support indicators (i.e., district provided a pacing guide or not; there was ongoing professional development in mathematics or not). Districts were coded as either providing higher levels of support or lower levels of support based on the relative presence or absence of the support indicators present. Ultimately, districts 1, 3, 4, 5, and 7 were considered in the high-support category while 2, 6, 8, 10, and 11 were considered low support.

The qualitative sub-study confirmed the importance of accountability for use and monitoring of the implementation, but this dimension of support was not analyzed as part of the district-level support variable included in the models. “Accountability and monitoring” seems to be a separate construct rather than a dimension of support. As part of the coding process of support, we found that some districts had low levels of accountability and monitoring but were high on dimensions of support; conversely, some districts had high accountability and low support. The two seemed to be operating as separate constructs and, therefore, in this paper we have limited our examination to the elements of support outlined in Study 1 excluding accountability and monitoring.

**Data Analysis Overview.** To examine the extent to which there was growth in the average mathematics achievement of schools’ fourth graders in the three time points prior to implementation of the mathematics instructional materials and in the three time points after implementation starting at the first year of implementation, we used HLM 7 software (Raudenbush, Bryk & Congdon, 2011) to test our models. Specifically, we used a two-level regression model in which measures at six time points were nested within schools; three
measures prior to implementation and three measures post implementation. An unconditional piece-wise model and multiple random coefficients models were formulated. Multilevel modeling provided a flexible modeling approach for examining the variability in growth across schools and for exploring predictors of that growth (Snijders & Bosker, 1999).

Results

Descriptive Statistics. Basic descriptive data are included in Figure 4 and Table 2, which show the mean z-scores, standard deviations, and sample size by district and implementation year. District-level averages for the schools in our sample indicate that all districts made gains from three years prior to implementation (M=0.323) to three years after implementation (M=0.508). Figure 5 shows the mean z-scores for the 10 districts in aggregate by implementation year and Table 3 includes the means and standard deviations for the full sample by implementation year. Mean z scores for the full study sample ranged from a low of 0.323 in Year-2 to a high of 0.508 in Year 3. Mean z-scores rose incrementally and fairly steadily, with the exception of a slight decline from Year-1 to Year 0.

Multilevel Piece-Wise Growth Curve Analysis. To allow for the possibility of different growth patterns pre- and post-implementation, we used a piece-wise approach in which time was separated into two segments. The three pre-implementation years were coded -2, -1, and 0, with 0 indicating the last year prior to the implementation of the new curriculum. The three post-implementation years were coded 1, 2, and 3, and overlapped with the pre-implementation time period by including 0. This coding scheme was used so that growth trajectories at different stages of the intervention could be estimated (Chou, Yang, Pentz & Hser, 2004; Wegner, Soumerai, Zhang, & Ross-Degnan, 2002). We coded the time segments in order to show the difference in the slopes between the two time periods (see Table 4).
The statistical models were constructed in stages. First, we constructed an unconditional, piece-wise growth model that included the two time variables as predictors of schools’ average fourth-grade mathematics achievement, represented by a z-score with a mean of 0 and a standard deviation of 1 (Model 1).

Model 1 took the following form:

Level 1 \[ \text{ZSCH}_{it} = \pi_{0i} + \pi_{1i} \times (\text{Time Segment 1}_{it}) + \pi_{2i} \times (\text{Time Segment 2}_{it}) + e_{it} \]

Level 2 \[ \pi_{0i} = \beta_{00i} + r_{0i} \]
\[ \pi_{1i} = \beta_{10i} + r_{1i} \]
\[ \pi_{2i} = \beta_{20i} + r_{2i} \]

The regression coefficients in the model allowed us to examine whether achievement was significantly different from zero in the year prior to implementation (\( \beta_{00} \)), whether there was growth in achievement in the three years prior to implementation (\( \beta_{10} \)), and whether there was growth in achievement during the implementation years (\( \beta_{20} \)). The random components in the model also allowed us to examine whether there was significant variation in achievement among schools in the year prior to implementation, and whether there was significant variation in growth among schools in the years prior to implementation and during the implementation years.

Second, to control for demographic differences among schools, the percentage of low-income students in each school was added to the school-level models as predictors of the level-1 variability (Model 2). This predictor was grand mean centered so that the intercept in the model was the predicted achievement for schools with an average percentage of low-income students. This model allowed us to examine whether the percentage of low-income students predicted differences in achievement among schools in the year prior to implementation (\( \beta_{001} \)), and whether the percentage of low-income students predicted growth in achievement in the three years prior to implementation (\( \beta_{101} \)) and during the implementation years (\( \beta_{201} \)).

Model 2 was as follows:
Level-1 \[ ZScore_{ti} = \pi_0 + \pi_1 \times (Time \ Segment 1) + \pi_2 \times (Time \ Segment 2) + e_{ti} \]
Level-2 \[
\pi_0 = \beta_{00} + \beta_{01} \times (Percent \ Low \ Income) + r_{0i} \\
\pi_{1i} = \beta_{10} + \beta_{11} \times (Percent \ Low \ Income) + r_{1i} \\
\pi_{2i} = \beta_{20} + \beta_{21} \times (Percent \ Low \ Income) + r_{2i}
\]

In the third and final model, we included a dichotomous variable at the school level to represent the support provided to schools by their districts during the implementation years (Model 3). The inclusion of this variable allowed us to examine whether, after controlling for the percentage of low-income students, the level of support provided by districts predicted differences in growth among schools during the implementation years.

Model 3 was as follows:

Level-1 \[ ZScore_{ti} = \pi_0 + \pi_1 \times (Time \ Segment 1) + \pi_2 \times (Time \ Segment 2) + e_{ti} \]
Level-2 \[
\pi_0 = \beta_{00} + \beta_{01} \times (Percent \ Low \ Income) + r_{0i} \\
\pi_{1i} = \beta_{10} + \beta_{11} \times (Percent \ Low \ Income) + r_{1i} \\
\pi_{2i} = \beta_{20} + \beta_{21} \times (Percent \ Low \ Income) + \beta_{22} \times (Support) + r_{2i}
\]

Table 5 shows the fixed and random effects for each model. The results for Model 1 indicate that there was no significant growth in achievement during the years prior to implementation; school achievement scores were predicted to increase by a non-significant 0.026 standard deviations (p = 0.066) between time points. In the year prior to implementation, the average predicted achievement score was 0.405 standard deviations, and this was statistically significantly different from zero (p < .001). During the implementation years, school achievement was predicted to increase by 0.043 standard deviations between each time point, and this growth was statistically significant (p < .001). The magnitudes of the coefficients indicate that the predicted growth in achievement between time points during the implementation years was almost twice that of the pre-implementation growth. The random components in Table 5 indicate that there was significant variation among schools’ growth trajectories in the years prior to implementation (p = .001) and during the implementation years (p < .001). Likewise,
there was significant variation among schools’ achievement during the year prior to implementation (p < .001).

The results from Model 2 indicate that the percentage of low-income students in a school was not significantly associated with the variability in schools’ growth either during the pre-implementation years (B_{01} = 0.012, p = .827) or in the year prior to the implementation of the intervention (B_{11} = -0.359, p = 0.059). During the implementation years, the percentage of low-income students in a school was significantly associated with the variability in schools’ growth (B_{21} = -0.085, p < .05); schools with higher percentages of low-income students were predicted to experience lower growth in achievement between time points. The residual variance components show that after controlling for the percentage of low-income students, significant variation remained among schools’ growth trajectories in the years prior to implementation (p = .001) and during the implementation years (p < .001). Likewise, after controlling for the percentage of low-income students, significant variation among schools’ achievement remained during the year prior to implementation (p < .001).

The fixed effects from Model 3 indicate that, after controlling for the percentage of low-income students in the school, the level of support provided by districts was a statistically significant predictor of the variability in growth during the implementation years. Specifically, schools that received higher levels of support from their districts were predicted to grow an additional 0.138 standard deviations between time points compared with schools that did not receive support (p < .001). However, the residual variance components in Table 5 indicate that district support explained no additional variance in growth above that explained by the percentage of low-income students alone, and that significant variation in growth remained among schools (p < .001).
Discussion

The results of the piece-wise growth curve analysis supported the main hypotheses of the study—that the implementation of instructional materials well-aligned to state standards and tests will improve student outcomes; and that implementation accompanied by high levels of district support will result in greater gains in average student mathematics achievement than implementations that are not as well-supported. Study 2 showed that the growth rate in school-aggregated mathematics achievement scores before implementation of the mathematics materials was not significant, in other words, school-aggregated fourth-grade achievement was not changing over time. This finding may, in part, be due to the fact that before implementation of the new mathematics materials, many of the districts in the study did not have a specific set of materials they were using across all grades in all elementary schools. For example, the materials that schools were using may have varied within schools, grades, or even classrooms. In addition, the districts in the study may have noticed a decline in their student achievement and thus were ready to implement a new strategy to improve achievement. It is also possible that the materials used in the schools pre-implementation were not well-aligned with the state assessments. However, during the three years of implementation of the new mathematics materials the growth rate was statistically significant. By using a piece-wise growth curve analysis we were able to predict these two different slopes before implementation and after implementation of the new mathematics materials. The good news is that this move of implementing new mathematics materials did improve student mathematics achievement over time. Additionally, this finding suggests that the implementation of a high-quality program may alone contribute to improvement in students’ mathematics achievement.
However, our data show that the effect of the implementation may be dampened in districts with a higher percentage of low-income students. Further analysis of the longitudinal data showed that schools with higher percentages of low-income students, while still experiencing growth in mathematics achievement post-implementation, experienced less growth than schools with lower percentages of low-income students. However, the percentage of low-income students at the school only explained a portion of the variation of schools’ achievement; there was still considerable variance unexplained. Therefore, the models suggest that additional factors may be contributing to the differences between schools.

In particular, we hypothesized that strong district-level support for implementation would lead to higher student achievement. The analysis showed that when controlling for the percentage of low-income students in a school, the district-level of support was a statistically significant predictor of growth in mathematics achievement post-implementation; in fact, schools in a highly supportive district were predicted to grow an additional 0.138 standard deviations each year compared with schools in districts with less support (p<0.001). However, the support from the district did not explain any additional variance in growth beyond that explained by the percentage of low income. This indicates that there is still significant variation between schools left to explain. Therefore, there may be additional factors that we have not measured that are contributing to the variation between schools in their mathematics growth. The models we have tested so far do not fully explain the differences between schools in their growth trajectories.

Our next step in the analysis is to examine the teacher survey data we have about supports at the school level and teachers’ use of the instructional materials during the third year of implementation. We hypothesize that the more support teachers receive, the more teachers are
using the materials (and, by extension, the more content they are teaching), leading to greater growth in student achievement. Future work will examine whether school-level support promotes mathematics achievement during the third year of implementation and whether teacher use of the materials mediates that relation while controlling for pre-implementation mathematics achievement.

Limitations and Future Directions

The present study has several limitations. First, our sample includes a select group of districts. We intentionally included districts that chose a district-level strategy centered on curriculum implementation of high-quality materials, and expected all teachers in all schools to use those materials. Participating districts may have provided more support for implementation than districts implementing other materials as both Everyday Mathematics and Investigations are widely understood in the mathematics education community to be programs that require substantial support to be used well. There is evidence in our interviews that district leaders across sites expected from the beginning of the implementation that these programs would require some support (though clearly, the districts varied in what they understood the necessary supports to be, and what they provided). In one of our states, districts had procured the materials through the use of waivers or other means because the programs were not fully supported on the state’s adoption list; as such, district leaders were highly motivated to get these programs, and their support of the programs may have been more than other districts.

Second, there are clear limitations to our assessment of district-level support in Study 2. In assessing the level of support provided by each district, we relied on retrospective interviews of district leaders conducted in the third year of implementation, in which they were asked to relate the history of the implementation and the supports they had provided. We may have
obtained a more valid and detailed measurement of those supports if those interviews had been conducted each year of the implementation. In addition, our district-level interview data tell us about the supports for implementation as planned and enacted at the district level; they do not tell us how those supports were experienced or received at the school or teacher level. In addition, in that data, we do have some indication that some districts intentionally varied the supports provided to each school, essentially performing triage to determine where the greatest needs for support lie. Our measure of support does not take this into account. Indeed, our measure of support does not allow for variation in the level of supports provided at the school level and does not take into account that schools are nested in districts. As we continue to analyze the school- and teacher-level data we have collected through surveys, we expect we will have a clearer understanding about how those supports varied from school to school within districts.

Third, we measured fourth-grade mathematics achievement using four different state tests. While we standardized the raw student scores around each state’s proficiency level, those proficiency levels are not all set the same when compared with the NAEP scores (as discussed in the Study 2 Methods section). In addition, the content covered and extent to which it is covered varies on the four tests. While the mathematics content standards for grade 4 are more similar than different across the four states, there are nonetheless differences between them and those differences are reflected in the content tested on the state exams. Unless we were to confine our study to a single state or ask the districts to administer a common test, this variation is currently unavoidable.

Finally, this study is correlational in nature as it is a quasi-experimental study with no random assignment or control group; the districts are being compared with themselves as part of a natural experiment. We are therefore not able to claim causation.
Further research on the implementation of mathematics materials is clearly needed in order to more fully explain why schools differ from each other in their growth of mathematics achievement. In addition, using a common metric to compare schools and districts across states would provide a better estimate of the differences in growth of student learning. For example, when the Common Core assessments are in full use, this study could be replicated.

Conclusions

Past research has shown that the materials teachers use in their classrooms relate to the content that is taught and what students learn. The present research suggests that districts’ support of the use of those instructional materials also matters, such that more support predicts growth in students’ mathematics achievement. In this paper, we presented a framework of district support for implementation validated through qualitative research (Study 1) that we then used to categorize 10 districts in our larger study (Study 2) into high-support and low-support districts. While further analysis remains, we have found that the level of district support significantly predicts school-aggregate fourth-grade students’ growth in mathematics achievement over the three years of implementation of new mathematics instructional materials. While it may seem self-evident that more support leads to better outcomes, many districts continue to struggle to provide support due to budget constraints and/or competing priorities. Further analysis of the dataset will suggest the kinds of supports that may be more effective and important for increasing teacher use of the materials and student outcomes.
References


Implications and applications for research and practice. *Psychology in the Schools, 45*(2), 158–176.


Figure 1. Relation Among the Constructs

Figure 1. This study explores the relation among the district-level support for implementation, school-level support for implementation, school-level use, and the effect on student outcomes. This paper focuses on the leg of this diagram labeled A—the relation between district-level support and school-level student outcomes as measured by the fourth-grade state mathematics test.
Figure 2. Mean State-Level Z-scores by Implementation Year (for the 10 districts in the sample)

Figure 2. Figure 2 shows the mean state-level z-scores by implementation year for the 10 districts in the sample.
Figure 3. Estimated NAEP Scale Equivalent Acores for Grade-4 State Mathematics Proficiency Levels, 2009 (NCES)
Figure 4. Mean Z-scores by District and Implementation Year
Figure 5. Mean Z-scores for Full Sample by Implementation Year
Table 1. Elementary Student Population Demographics by District

<table>
<thead>
<tr>
<th></th>
<th>District 1</th>
<th>District 2</th>
<th>District 3</th>
<th>District 4</th>
<th>District 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Student population</strong></td>
<td>23,545</td>
<td>3,882</td>
<td>5,855</td>
<td>6,484</td>
<td>26,159</td>
</tr>
<tr>
<td><strong>Elementary Student population</strong></td>
<td>11,732</td>
<td>1,717</td>
<td>2,177</td>
<td>3,298</td>
<td>12,852</td>
</tr>
<tr>
<td><strong>School sample</strong></td>
<td>21</td>
<td>3</td>
<td>5</td>
<td>21</td>
<td>24</td>
</tr>
<tr>
<td><strong>4th grade sample</strong></td>
<td>1486</td>
<td>248</td>
<td>443</td>
<td>403</td>
<td>1810</td>
</tr>
<tr>
<td><strong>M SD M SD M SD M SD M SD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>% free/reduced lunch</strong></td>
<td>60.1</td>
<td>24.1</td>
<td>18.0</td>
<td>10.6</td>
<td>30.7</td>
</tr>
<tr>
<td><strong>% White</strong></td>
<td>46.1</td>
<td>19.5</td>
<td>47.2</td>
<td>12.5</td>
<td>73.8</td>
</tr>
<tr>
<td><strong>% Black</strong></td>
<td>6.2</td>
<td>4.9</td>
<td>4.1</td>
<td>2.5</td>
<td>3.4</td>
</tr>
<tr>
<td><strong>% Hispanic</strong></td>
<td>39.6</td>
<td>19.7</td>
<td>9.8</td>
<td>2.5</td>
<td>13.4</td>
</tr>
<tr>
<td><strong>% Asian</strong></td>
<td>4.9</td>
<td>4.3</td>
<td>35.7</td>
<td>6.4</td>
<td>4.1</td>
</tr>
<tr>
<td><strong>% Multi-race</strong></td>
<td>2.4</td>
<td>1.4</td>
<td>1.8</td>
<td>1.3</td>
<td>5.0</td>
</tr>
<tr>
<td><strong>% American Indian/Alaskan Native</strong></td>
<td>0.6</td>
<td>0.5</td>
<td>0.8</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td><strong>% Hawaiian Native/Pacific Islander</strong></td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>District 6</th>
<th>District 7</th>
<th>District 8</th>
<th>District 9</th>
<th>District 11</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Student population</strong></td>
<td>8,126</td>
<td>3,452</td>
<td>7,150</td>
<td>17,509</td>
<td>29,842</td>
</tr>
<tr>
<td><strong>Elementary Student population</strong></td>
<td>3,626</td>
<td>1,215</td>
<td>3,294</td>
<td>7,973</td>
<td>18,173</td>
</tr>
<tr>
<td><strong>School sample</strong></td>
<td>8</td>
<td>4</td>
<td>7</td>
<td>17</td>
<td>36</td>
</tr>
<tr>
<td><strong>4th grade sample</strong></td>
<td>653</td>
<td>262</td>
<td>457</td>
<td>1152</td>
<td>2125</td>
</tr>
<tr>
<td><strong>M SD M SD M SD M SD M SD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>% free/reduced lunch</strong></td>
<td>28.4</td>
<td>12.3</td>
<td>15.8</td>
<td>4.3</td>
<td>45.2</td>
</tr>
<tr>
<td><strong>% White</strong></td>
<td>89.3</td>
<td>5.5</td>
<td>92.5</td>
<td>2.7</td>
<td>58.2</td>
</tr>
<tr>
<td><strong>% Black</strong></td>
<td>2.1</td>
<td>1.7</td>
<td>1.7</td>
<td>0.8</td>
<td>7.0</td>
</tr>
<tr>
<td><strong>% Hispanic</strong></td>
<td>2.9</td>
<td>2.9</td>
<td>1.7</td>
<td>0.7</td>
<td>8.5</td>
</tr>
<tr>
<td><strong>% Asian</strong></td>
<td>1.6</td>
<td>0.8</td>
<td>3.7</td>
<td>2.2</td>
<td>11.0</td>
</tr>
<tr>
<td><strong>% Multi-race</strong></td>
<td>3.8</td>
<td>1.3</td>
<td>0.0</td>
<td>0.0</td>
<td>14.0</td>
</tr>
<tr>
<td><strong>% American Indian/Alaskan Native</strong></td>
<td>0.2</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.9</td>
</tr>
<tr>
<td><strong>% Hawaiian Native/Pacific Islander</strong></td>
<td>0.2</td>
<td>0.2</td>
<td>0.0</td>
<td>0.4</td>
<td>0.3</td>
</tr>
</tbody>
</table>
Table 2. Descriptive Statistics of School-level Z-scores by District and Implementation Year

<table>
<thead>
<tr>
<th>District</th>
<th>Year -2</th>
<th>Year -1</th>
<th>Year 0</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>1</td>
<td>.94</td>
<td>.419</td>
<td>1441</td>
<td>.92</td>
<td>.340</td>
<td>1455</td>
</tr>
<tr>
<td>2</td>
<td>-.05</td>
<td>.196</td>
<td>235</td>
<td>.41</td>
<td>.135</td>
<td>237</td>
</tr>
<tr>
<td>3</td>
<td>.183</td>
<td>-</td>
<td>.231</td>
<td>-.31</td>
<td>.183</td>
<td>.448</td>
</tr>
<tr>
<td>4</td>
<td>.60</td>
<td>.404</td>
<td>389</td>
<td>.46</td>
<td>.351</td>
<td>392</td>
</tr>
<tr>
<td>5</td>
<td>1.06</td>
<td>.282</td>
<td>1669</td>
<td>.95</td>
<td>.311</td>
<td>1710</td>
</tr>
<tr>
<td>7</td>
<td>1.18</td>
<td>.377</td>
<td>253</td>
<td>1.12</td>
<td>.401</td>
<td>237</td>
</tr>
<tr>
<td>8</td>
<td>0.03</td>
<td>.348</td>
<td>471</td>
<td>.24</td>
<td>.331</td>
<td>431</td>
</tr>
<tr>
<td>10</td>
<td>0.18</td>
<td>.385</td>
<td>1236</td>
<td>.35</td>
<td>.311</td>
<td>1225</td>
</tr>
</tbody>
</table>

Table 2. District-level averages for the schools in our sample indicate that all districts made gains from Year-2 (two years prior to implementation) to Year3 (three years after implementation).
<table>
<thead>
<tr>
<th>Year</th>
<th>Mean</th>
<th>N</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year-2</td>
<td>.323</td>
<td>130</td>
<td>.66672</td>
</tr>
<tr>
<td>Year-1</td>
<td>.439</td>
<td>130</td>
<td>.67660</td>
</tr>
<tr>
<td>Year0</td>
<td>.339</td>
<td>130</td>
<td>.61039</td>
</tr>
<tr>
<td>Year1</td>
<td>.493</td>
<td>131</td>
<td>.62343</td>
</tr>
<tr>
<td>Year2</td>
<td>.519</td>
<td>131</td>
<td>.70742</td>
</tr>
<tr>
<td>Year3</td>
<td>.508</td>
<td>131</td>
<td>.68752</td>
</tr>
</tbody>
</table>

*Table 3. Mean School-level Z-scores for Full Sample by Implementation Year*

Mean z-scores for the full study sample ranged from a low of .323 in Year-2 to a high of .508 in Year3. Mean z-scores rose incrementally and fairly steadily, with the exception of a slight decline from Year-1 to Year0.
Table 4. Time Segment Coding for Piece-wise Linear Model

<table>
<thead>
<tr>
<th>Time Segment 1 - growth rate prior to implementation</th>
<th>Year-2</th>
<th>Year-1</th>
<th>Year0</th>
<th>Year1</th>
<th>Year2</th>
<th>Year3</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

| Time Segment 2 - growth rate after implementation   | 0      | 0      | 0     | 1     | 2     | 3     |
Table 5: Fixed Effects and Random Effects of the Multilevel Models

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Effects</td>
<td>Coeff (se)</td>
<td>p</td>
<td>Coeff (se)</td>
</tr>
<tr>
<td>Pre-Implementation Growth</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, B_{00}</td>
<td>0.026 (.01)</td>
<td>.065</td>
<td>0.026 (.01)</td>
</tr>
<tr>
<td>Percent Low Income, B_{01}</td>
<td>0.012 (.05)</td>
<td>.827</td>
<td>0.012 (.05)</td>
</tr>
<tr>
<td>Year Prior to Implementation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, B_{10}</td>
<td>0.405 (.05)</td>
<td>&lt;.001</td>
<td>0.405 (.05)</td>
</tr>
<tr>
<td>Percent Low Income, B_{11}</td>
<td>-0.359 (.19)</td>
<td>.059</td>
<td>-0.359 (.19)</td>
</tr>
<tr>
<td>Growth during Implementation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, B_{20}</td>
<td>0.042 (.01)</td>
<td>&lt;.001</td>
<td>0.042 (.01)</td>
</tr>
<tr>
<td>Percent Low Income, B_{21}</td>
<td>-0.085 (.04)</td>
<td>.032</td>
<td>-0.104 (.04)</td>
</tr>
<tr>
<td>Support, B_{22}</td>
<td>0.138 (.02)</td>
<td></td>
<td>0.138 (.02)</td>
</tr>
<tr>
<td>Random Effects</td>
<td>Available Variance</td>
<td>p</td>
<td>Residual Variance</td>
</tr>
<tr>
<td>Variance in growth during pre-implementation years</td>
<td>0.007</td>
<td>.001</td>
<td>0.007</td>
</tr>
<tr>
<td>Variance in achievement during the year prior to implementation</td>
<td>0.347</td>
<td>&lt;.001</td>
<td>0.341</td>
</tr>
<tr>
<td>Variance in growth during implementation years</td>
<td>0.006</td>
<td>&lt;.001</td>
<td>0.005</td>
</tr>
</tbody>
</table>