Coherent Implementation of Mathematics Instructional Materials:

A Multilevel Mixed Methods Study

Deborah B. Spencer

Education Development Center, Inc., Waltham, Massachusetts

John W. Creswell

University of Nebraska-Lincoln

Kristen E. Reed, Jessica M. Young, June Mark

Education Development Center, Inc., Waltham, Massachusetts

Deborah B. Spencer, Learning and Teaching Division, Education Development Center, Inc.; John W. Creswell, College of Education and Human Sciences, University of Nebraska, Lincoln; Kristen E. Reed, Learning and Teaching Division, Education Development Center, Inc.; Jessica M. Young, Learning and Teaching Division, Education Development Center, Inc.; June Mark, Learning and Teaching Division, Education Development Center, Inc.

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: Katherine Schwinden, Julie Zeringue, Deborah Rosenfeld, Mari Halladay, Jess Gropen, Kathryn Chiappinelli, and Victor Mateas. And thanks to Kerry Ouellet who helped with manuscript preparation.

Correspondence concerning this article should be addressed to Kristen E. Reed, Education Development Center, Inc., 43 Foundry Ave., Waltham, MA 02453.
Email: kreed@edc.org
Abstract

Multilevel assessments in education settings typically include five levels: states, school districts, schools, classrooms or teachers, and students. In the mixed methods literature, multilevel designs remain relatively unexplored as a specific type of approach. This study advances the mixed methods literature by addressing procedures and challenges that arise in this form of a complicated design. We discuss the process of conducting this multilevel mixed methods study and focus on seven mixed methods procedures used in this study: the use of a visual display to represent the multilevel design; the identification of the points of interface of the quantitative and qualitative datasets, the use of parallel questions and constructs, the use of joint displays; the approach to resolving discrepancies between our datasets; the composition of the project team, and the plan for publishing findings. We advance specific recommendations for the use of this type of mixed methods design.
The prevalence of mixed methods research among the social, behavioral, and health sciences is now widely recognized (Onwuegbuzie, 2012). This prevalence is seen in numerous empirical and methodological articles, the establishment of key journals devoted to mixed methods research, numerous books available with different discipline orientations, and initiatives of large funding agencies to establish criteria for assessing the quality of mixed methods proposals and assist reviewers in evaluating projects (see, the National Institutes of Health “Best Practices,” Creswell, Klassen, Plano Clark, & Smith, 2011). Within this unfolding mixed methods field, no topic has been more widely explored than the types of designs. As many as 13 different classifications—developed by different authors in diverse fields—have been noted (Creswell & Plano Clark, 2011). This discussion has led to the use of diagrams of designs, specific procedures for indicating the flow of activities, a notation system to provide symbols in the diagrams, the identification of methodological issues and challenges that arise in the implementation of designs, and knowledge of idiosyncratic features of designs within fields. For example, in the humanitarian fields of crisis interventions, the emphasis on designs has been to link to program impact (Bolton, et al., 2007). In distance learning, Ivankova and Stick (2007) argued that following a survey with interviews and building case studies would be the preferred design. In the health sciences, embedding the randomized controlled trial with qualitative data—either before the intervention trial or after to follow up with results—has been widely discussed as an important design to use.

In education, complex issues require attention to multiple levels of involvement that span from the district, to the school, to the classroom, and on to the individual student, and studying
these issues may require yet another type of mixed methods design—a multilevel design. Unfortunately, we know little about how these designs work, aside from the example of a large-scale state education evaluation project reported by Tashakkori and Teddlie (1998). We know even less about the challenges these designs present from the standpoint of team involvement, bridging different levels of data and analysis, and incorporating multiple complex databases. The study described in this paper intends to advance a multilevel mixed methods study design, and to assess both the benefits and the challenges of using this design within a mixed methods framework. We define mixed methods research as the use of multiple methods of data (quantitative and qualitative) that are responsive to both quantitative and qualitative research questions. Further, these databases are integrated, or combined, in order to draw on the strengths of both forms of data. A multilevel mixed methods design is then one option for a researcher when working across an organizational structure composed of multiple levels of input into a problem.

**Coherent Implementation of Mathematics Instructional Materials Study**

In this paper, we illustrate a multilevel mixed methods design using a recent study of school districts’ implementation of the elementary mathematics materials *Everyday Mathematics* and *Investigations in Number, Data, and Space*. Our study is an NSF-funded, four-year effort investigating the effects of a district-level strategy for instructional improvement, centered on the consistent use of mathematics instructional materials in the elementary grades. Our 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 (as shown in Figure 1).

To address our key research questions, the study has employed both qualitative and quantitative data. Those questions are as follows:
1. For each of the 12 districts in our study, what is the district-level support measured quantitatively and assessed qualitatively? Do these data converge? Are the perceptions of district-level support consistent across the district, school, and teacher levels?

2. For each of the schools in our study, what is the school-level support measured quantitatively and assessed qualitatively for our interview subsample schools? For that subsample, do these data converge?

3. How do district- and school-level support relate to changes in school-level student scores? What factors moderate the relation between district- and school-level support and student outcomes (e.g., percent free or reduced-price lunch, size of school)? Does school-level use mediate the relation between district- and school-level support and student outcomes?

Our overall hypothesis, building on the literature in the field and the results of our prior work (Mark, Spencer, Zeringue, & Schwinden, 2010), was that school districts that had greater coherence in their support for curriculum implementation would see stronger student outcomes. We expected that district-level support for implementation, when paired with support for implementation at the school level, would result in greater use of the instructional materials at the classroom level. That higher level of use of the materials—materials that were aligned with state standards and measures and consistent across classrooms and grade levels—might have resulted in increases in student performance on the mathematics portions of state tests. Conversely, we expected that a low level of support from the district and the school for implementation would be correlated with a low level of classroom use and, therefore less likely to have a positive effect on student outcomes.
Our sample consisted of 12 districts in five states in the U.S. and included over 2,000 elementary teachers from 153 schools. Our sample was divided into two cohorts of six districts each. Each participating district met several criteria, including:

- **Curriculum use.** Each district was using one of two elementary mathematics curricula—*Everyday Mathematics* or *Investigations in Number, Data, and Space*—in all of their elementary schools. All mathematics teachers in grades 1–4 were included in the study.

- **Adoption year.** Each district had to be in its third year of use of the curriculum, as we expected the effects of implementation on student learning to be more visible after three years; this period of time allowed teachers to become adept in the use of the materials and students to experience the program in multiple grades. Our Cohort 1 districts adopted the curricula in 2008–09; our Cohort 2 districts, in 2009–10.

- **District size.** We limited district size to 3,000–35,000 students, because we believed that district-level efforts to support implementation of the materials would be more visible and articulated in districts large enough to have had to establish a strategy for implementation across multiple elementary schools. We ruled out very large districts that might require an additional layer of bureaucracy in order to manage curriculum and instruction.

- **Limitations on state context.** Our outcome data was gathered from state mathematics test results for fourth graders, aggregated to the school level. The study employed a short interrupted time series design, using six years of outcome data—three prior to adoption of new materials and three following. Therefore, participating districts had to be in states that had stable state standards and a stable state test system from 2005–2012. We aimed
to group districts in as few states as possible, and the resulting sample represented five states.

We collected quantitative and qualitative data in our 12 districts to address our key research questions. Methods included surveys ($N=2,172$ teachers, 153 principals), school-level interviews ($N=111$), district-level interviews ($N=26$), and 12 site visits. We surveyed classroom teachers four times over the school year about their use of *Investigations* or *Everyday Mathematics*, and their perceptions of school- and district-level support. We surveyed principals twice about school- and district-level support. We interviewed district-level administrators (mathematics administrators and supervisors of principals) and collected district-level documents—primarily pacing guides and other artifacts of curriculum planning and support. In a subsample of 30 schools, we interviewed two to three teachers and the principal or mathematics specialist. As we are now in our fourth year, we can reflect back over the process of conducting the investigation and identify key elements of mixed methods research that informed our procedures. We can also, now, better assess the challenges entailed in conducting a multilevel mixed methods investigation.

**Mixed Methods Procedures Used in the Study**

There are seven mixed methods aspects of this investigation that may be of use to others conducting a multilevel mixed methods project: the use of a visual display to represent our multilevel design, the identification of the points of interface of the quantitative and qualitative datasets at different levels of our design, the use of parallel questions and constructs, the use of joint displays, our approach to resolving discrepancies in our datasets, the composition of the project team, and our plan for publishing findings from our study.

**A Visual Display to Represent a Multilevel Design**
A visual display of the design presents the overall plan for the multi-level mixed methods study and it can show the major constructs assessed, the types of quantitative and qualitative data collected, and the places where the two data sources interface. From the beginning of this study, we have worked to clarify our representation of the data sources that would inform each of our constructs. Using a visual display of this design proved to be helpful internally—serving as a reminder to our team about the many aspects of our study and helping us clarify how the data would be integrated at the points of interface between the quantitative and qualitative datasets—and in communicating our design to partners, advisors, and others interested in our study. The display (see Figure 2) represents our four key constructs, our quantitative sources of data for each construct, our qualitative sources of data for each construct, and the points of interface between the two datasets for each construct. Of particular interest to some readers may be the multilevel nature of the display—represented by the four constructs in the left-hand Constructs column, and the associated data sources and points of interface for each. Because of the multilevel nature of our study, we have multiple points of interface between the qualitative and quantitative data, depending on the construct and level of the data.

**Points of Interface of the Quantitative and Qualitative Datasets**

A point of interface (Morse & Niehaus, 2009) is the place in the plan or design where the two databases intersect. As you can see in Figure 2, our visual display makes clear that there were three points of interface between our quantitative and qualitative datasets, centered on three of our constructs: district-level support, school-level support, and teacher use of the instructional materials. (Our fourth construct, student outcomes, was measured using only quantitative data, which were scores for fourth graders on the state mathematics test, aggregated to the school level.) For school-level support and teacher use, we employed two techniques for integration of
the data: we used both convergent and explanatory analysis. For the district-level of support, we employed only merged analysis. Below we describe these points of interface.

**District-level support: Fixed sample selection and merged analysis.** For our district-level support construct, our use of mixed methods has allowed us to consider multiple perspectives on support, from across levels of each participating school district—attending to both the intent and provision of support as described by district leaders in interviews, and the experience and effects of those supports as reported by principals and teachers on surveys. For each of the 12 districts in our sample, we interviewed district-level administrators (mathematics administrators and supervisors of principals), conducted a site visit, and collected district-level documents (primarily pacing guides and other artifacts of curriculum planning and support). We asked in our interviews about their experiences supporting curriculum implementation, including basic supports provided to teachers, like materials and instructional time needed; their efforts to build the capacity of teachers and administrators to use the materials well; and their efforts to monitor use of the materials and establish accountability for use. Our quantitative data on this construct came from items on principal and teacher surveys about their perspectives on types of support provided by the district. While the quantitative and qualitative data were collected from different levels, our use of parallel questions allowed us to merge the results and compare perspectives across levels. Our primary tool for merging these datasets has been the use of multilevel joint displays, in which our quantitative and qualitative data were summarized and compared along dimensions of our district-level support construct. This merged analysis provided a more complete picture of district-level support than either would provide alone.

**School-level support and teacher use of the materials: Sequential sample selection with both convergent and explanatory analysis.** To enhance our understanding of our school-
level support and teacher use constructs, we interviewed teachers and principals in a sub-sample of 40 of our 153 participating schools. We introduced these subsample interviews into our design for two reasons: explanation and convergent validity.

The first purpose was explanation, at the thematic or phenomena level. We wanted to know how and why teachers used the materials, how they were supported in using the materials, and what effects support had on use. We used an explanatory sequential design (Creswell & Plano Clark, 2011) in this aspect of our study: we wanted a sample of schools in the qualitative sub-study that reflected a range of reported use of materials at the school level in order to achieve comparability across different types of cases on our dimensions of interest (Teddlie & Yu, 2007), and we used our initial survey results on use to identify the sample.

To identify a range of schools for the interview subsample, we relied on our early survey results on use. To efficiently identify each school’s level of use, we analyzed the responses to just two questions on the second teacher survey (of four). Question 1 asked teachers to select a statement that best describes how they use their mathematics instructional materials, and question 2 asked them to describe how most teachers in their schools use the materials. We aggregated the data to the school level and calculated a mean “level-of-use” score for each school along with the standard deviation. Each school was then ranked on their level of use within their district and within the overall sample. Based on their rankings, and distributed across the 12 districts, we chose a set of high use schools, a set of medium use schools, and a set of low use schools. We also took into account the standard deviation of each of the schools, where a higher standard deviation would indicate that teachers were reporting differing levels of use. This was particularly important for the medium level-of-use schools, in which it was possible to have
a bimodal split of high users and low users within the same school. We therefore chose
additional medium level-of-use schools in order to explore these differences more fully.

Once our sample was established, we proceeded with conducting our interviews and
analyzing the data thematically. We taped, transcribed, and summarized a total of 111 interviews
with teachers and principals at 40 different schools (the number of interviews per school varied,
we aimed for three teachers and a principal but we did not always achieve those numbers). We
coded all interviews using an a priori code list that we derived directly from our dimensions and
constructs, with a few additional themes included (for example, grade-level differences) and an
“other” code included that allowed us to note aspects of support and use that were not addressed
by our model. Currently, we have coded all the interviews, and we have begun to analyze these
code categories, looking for patterns, subthemes, and outliers that explain support and use.

Our second purpose was convergent validity—we aimed to increase our confidence in our
survey data. Our study relies on self-report teacher surveys and does not include any classroom
observation. We believed there would be several validity threats associated with our reliance on
survey instruments to measure our constructs of teachers’ use of Everyday Mathematics and
Investigations. First, we believed we might encounter response bias—would teachers with higher
use of the instructional materials be more likely to respond to our surveys because of their
investment in the materials themselves? Second, would teachers in schools with higher levels of
accountability skew their responses to provide a more positive picture of use? And finally, since
we were relying on self-report survey mechanisms to determine level of use and degree of
support provided at particular times of the year, would that data provide an accurate portrait of
use across the year? We considered similar threats to our school-level support data—would
teachers in more supportive schools be more likely to respond? Would they feel pressure to over-report the level of support they received?

Using the same interview data employed in the explanatory analysis above, we created qualitative profiles of school-level support and teacher use of the materials for the 23 schools for which we conducted three or four interviews. The subsample interviews were designed to collect additional information on teacher use of the materials and school-level support, using questions parallel to our survey questions. The qualitative profiles were then aligned with the quantitative data that we gathered from the same schools and directly compared on parallel constructs, using a joint display for each school (discussed below in Joint Displays). This comparison allowed us to ask, “To what extent do the quantitative and qualitative data converge? How and why? And how should this deeper understanding of the construct influence our quantitative analysis?” Our purpose for doing this comparison was to strengthen our analysis of the quantitative data that we had for each of the 153 schools in our sample overall.

**Use of Parallel Constructs and Questions**

Parallel constructs mean that similar questions are asked on the instruments and protocols in both the quantitative and qualitative database (Creswell & Plano Clark, 2011). Our theoretical model of a coherent district implementation contains three main constructs: district-level support, school-level support, and school-level use. To take advantage of the power of mixed methods strategies (e.g., the use of a joint matrix), we had to address how these constructs could be operationalized both qualitatively and quantitatively—as we designed parallel questions on survey and interview instruments, and as we planned which qualitative and quantitative analysis techniques to employ. The use of parallel constructs and questions was critical to our design in
that it allowed us to compare our databases directly in our merged, or convergent, analyses (Creswell & Plano Clark, 2011).

Below is a description of our three constructs at the district, school, and teacher levels:

- **District-level support.** We identified three dimensions of district-level support for implementation: *basic components* included sufficient instructional time and materials, and the provision of a pacing guide; *building knowledge and capacity* included professional development to support teachers in understanding and using the materials effectively, and principals to support and supervise teachers; *making it happen* included districts cultivating a consistent vision for mathematics instruction across levels, and holding principals accountable for the implementation of the materials in their schools.

- **School-level support.** We had three dimensions of school-level support for implementation that largely paralleled the district-level support dimensions: *basic elements* (e.g., sufficient instructional time and materials, pacing guides), *ongoing support and professional development*, and *accountability and monitoring*.

- **School level of use.** We determined the school level of use by considering the *regularity* with which the teachers use the materials, the *dose* of the materials students receive, and teachers’ consistency with the *mathematical storyline*.

The constructs were developed from years of qualitative research and hands-on work in many school districts (e.g., Mark, Spencer, Zeringue, & Schwinden, 2010; Spencer, Mark, Zeringue, & Schwinden, 2010; Goldsmith, Mark, & Kantrov, 2000). Transforming these constructs and dimensions into quantitative terms presented several challenges. First, we had to tease apart the complexities of each. For example, we separated school-level support into nine dimensions, including pacing guide, time for instruction, materials needed, and professional
development. Some of these dimensions contained further subcomponents; for example, the
dimension of professional development was further broken into professional development
workshops, grade-level planning time for mathematics, support from the mathematics specialist
in the building, etc. We developed survey questions to address these dimensions; we then put the
same level of work into developing our qualitative interview protocols to make sure we asked
about the dimensions and subcomponents addressed in the survey. However, given that teachers
and principals had a limited amount of time for an interview, we sometimes were not able to
cover all of the questions in the depth that we would have liked (Table 1).

Ensuring that our constructs and questions were parallel across our quantitative and
qualitative databases was been a critical step that then allowed us to use techniques that merge
the two and consider our level of convergence.

**Use of Joint Displays**

A joint display arrays the quantitative and qualitative results together in a table or a graph
so that the two databases can be compared and potential differences addressed (see, for example,
the table in Lee & Greene, 2007). We used joint displays as an analysis tool at two levels in our
design: at the district level for all 12 participating districts and at the school level for the 23
subsample schools with three or four interviews per school. At the district level, the joint
displays were used to merge the multilevel datasets and compare the perspectives of teachers,
principals, and district leaders across our district-level support dimensions. At the school level,
for our constructs of school-level support and use, the displays are used for explanation and to
validate our survey data. Below, we highlight the school-level joint displays.

As explained above, these school-level joint displays were created by aligning the
qualitative data we collected from our 23 interview sample schools with the quantitative data
from those same schools, generating 23 independent displays. We had over 1,000 quantitative variables in our quantitative database so, clearly, not all items could be included in these joint displays. To determine which items to include, we referred to our “construct-variable-item” charts, which identified which items corresponded to which constructs and dimensions. We chose the quantitative items that were parallel to the interview questions, and that gave us good data to compare. For several of these items (e.g., use of the various components of the materials), we had created composite variables that aggregate teachers’ responses for use in our quantitative analysis. In these joint displays, we chose not to use these aggregate variables—we wanted to use the data in its “less digested” form so that it would remain similar to the qualitative data (which could not be aggregated in the same way) and would retain the parallelism between the two datasets.

Table 2 is an example of the joint display for one interview sample school on the school-level support dimension of pacing guides, a dimension of school-level support. There were three survey items that pertained to pacing guides: (a) Does your district provide mathematics pacing guides and/or instructional charts that offer guidelines about which unit to teach and when to teach it? (b) In general, are you able to keep pace with the instructional chart/pacing guide that your district provides for the curriculum? and (c) How do you use the mathematics pacing guides and/or instructional charts provided by the district? The descriptive quantitative information for each question is in the left column, the summary of what we learned from the qualitative data about each of these questions is in the middle, and to the right is a comparison of the two datasets on each of these items. In Table 2, the data largely confirm each other. In the next section, we discuss our procedures for when the data were not confirming.

**Resolving Differences in Joint Displays**
When comparing qualitative and quantitative datasets, the data may confirm, disconfirm, or contradict each other. Some approaches to resolving such discrepancies are discussed in the literature, including acknowledging and treating the methods as fundamentally different, exploring further the methodological rigor of the components, exploring dataset comparability, and collecting further data to resolve differences (Farquhar, Ewing, & Booth, 2011). In the example (Table 2) of a joint display for the dimension “pacing guide” for one of the 23 subsample schools, the qualitative and quantitative data are largely confirming. However, this was not always the case, and we were particularly interested in investigating points where the data did not converge. These points could give us important insights into understanding how our constructs function in our subsample schools and the potential limitations of our quantitative dataset.

In considering the joint display for each subsample school, we looked at each dimension of school-level support and teacher use—for example, *professional development,* or *regularity of use*—and compared the data we had summarized quantitatively with what we had summarized qualitatively. We then completed what we called a “convergence record” for each school. In this record, we asked ourselves two questions about each dimension (or sub-dimension): (a) to what degree were the data parallel (on a 5-point scale from strongly agree to strongly disagree) and (b) to what degree did the data converge (on a 5-point scale from confirming to disconfirming). If we were not sure that the data are parallel, we could not assess the level of convergence. If we agreed or strongly agreed that the data are parallel, we could then respond to the question about the level of convergence of the data. If we then determined that the data for the school for a dimension was divergent on a particular dimension (or in some cases, inconclusive), we noted why. Importantly, we also noted if there were issues that should be considered in the broader
quantitative analysis. We then aggregated those joint matrices across the sub-sample of 23 schools, by dimension, to determine whether our qualitative data was or was not largely confirming our quantitative data, and what the source of any discrepancies might have been. Figure 3 details the steps we took in this analysis of our school-level joint displays, as we checked for parallel data, checked for convergence, and reanalyzed our datasets as needed to ensure parallel datasets and resolve discrepancies where possible.

We have found it critical to be careful about whether our joint displays reflect truly parallel data in our quantitative and qualitative summaries. For example, for one of our schools, we discovered that—for a few dimensions—the summarized qualitative data entered into the joint display were not parallel with the quantitative data and, furthermore, did not represent all that we knew about that dimension anecdotally. Therefore, we revisited our interview transcripts and reanalyzed them targeting the key themes that aligned with the quantitative results. This discrepancy between the datasets occurred because there were some quantitative items that dug deeply into particular aspects of our dimensions, such as the use of specific components of the materials (e.g., the games or daily math message)—and we did not always capture that level of detail in our qualitative summaries. Fortunately, it was easy to go back and gather that data from our transcripts and memos. Once we had parallel data for these dimensions, we were able to go back and assess the level of convergence in our convergence record.

Cases where the data were parallel but divergent gave us the opportunity to probe more deeply to understand our dimensions and what was occurring in each school. For the school-level joint displays, we had more quantitative data than qualitative data. The quantitative sample includes all teachers grades 1 to 4, and the response rate was generally 70% or better in each school, so our survey data are largely representative of each building. On the other hand, the
qualitative sample is only a subset of people in the school, typically three teachers and the principal. Our quantitative data is more representative of the school as a whole, but the qualitative data provides a nuanced description of the teachers’ and principal’s experiences. One school provides a particularly interesting example of an instance in which we had to go back to the qualitative data to get more information and then perform some additional analysis with the quantitative data to confirm (or disconfirm) our hypothesis. The quantitative data showed that there was variation in the number of teachers reporting that they participate in grade-level planning meetings—50% reported weekly, 42% reported monthly, and one teacher reported not at all. The qualitative data largely confirmed this finding that there are weekly or monthly meetings. We hypothesized that the fourth-grade teacher met less frequently in grade-level planning or not at all because in grades 4 and 5, the core subjects are departmentalized and only one teacher teaches mathematics.

However, when we took a closer look at the survey responses, we discovered that two of the first-grade teachers reported meeting weekly but two reported meeting monthly; two of the three second-grade teachers reported meeting weekly, but the third second-grade teacher reported not meeting at all; the third-grade team was more consistent and all reported meeting monthly. Interestingly, the fourth-grade teacher, who is departmentalized, reported meeting weekly. This more in-depth quantitative analysis disconfirms our hypothesis and makes us question the dynamics in the school and wonder if some grade-level teams and some teachers are more cohesive and tend to collaborate more than others.

**Composition of the Project Team**

Much attention recently in the mixed methods literature addresses the role, composition, and challenges of academic teams conducting mixed methods research (O’Cathain, Murphy, &
Nicholl, 2008). In our project, the composition and operation of our project team has been critical to our ability to carry out this complicated multilevel mixed methods design. Our study is intentionally staffed with professionals who have a range of analytic backgrounds—four with qualitative experience, two with quantitative experience. The qualitative team includes two former elementary teachers with a deep understanding of the school contexts and curricula we are studying. The other two researchers with a background in qualitative methods—our principal investigators—bring a thorough understanding of mathematics teaching and learning across grades as well as expertise in the design and implementation of mathematics instructional materials, K–12. We also have two team members with experience in quantitative methods; one, a cognitive scientist, and another, a developmental and educational psychologist—both of whom have led previous studies that were primarily quantitative. These researchers have a rich understanding of quantitative methods, having led large-scale, quasi-experimental and evaluation studies utilizing survey data.

From the beginning of our project—even in the proposal stage—we took advantage of the knowledge base of the quantitative and qualitative team members, so that these methodologies were integrated in our overall design, and each team member benefited from understanding the approach of the others. Our survey instruments and interview protocols were created and reviewed jointly to ensure that the questions we asked were parallel and still authentic to each method. In fact, our qualitative team members were integrally involved in writing and revising survey questions, piloting those questions, reviewing responses, and tracking response rates in our schools and districts, whereas our quantitative members also reviewed the questions for variation in survey responses and for analysis considerations. Moreover, our quantitative team members interviewed teachers and principals at our subsample
schools, and participated in discussions as we reviewed our qualitative profiles of each subsample school. A qualitative and a quantitative team member also worked together to conduct a preliminary analysis of initial survey data in order to select the interview subsample of schools. In addition, our members with qualitative expertise have participated in quantitative methods trainings that led to the creation of indices and composites that have allowed us to reduce our quantitative data.

Arguably, some of these project activities could have been carried out solely by a team member with expertise in that methodology; however, we chose to intentionally “mix” staff expertise on these tasks. This approach has proved invaluable to us in our coordination as a team, and in the resulting understandings that we have been able to carry into our truly “joint” work: the creation of survey and interview instruments that pursue parallel constructs, the establishment of our qualitative interview subsample using our initial survey data, and the comparison of our quantitative and qualitative data on parallel questions in our school-level and multilevel joint displays.

We had several advantages that facilitated operating this way. One was that our home institution, the Education Development Center, Inc. (EDC), is a project-based organization that by its nature encourages and even requires us to reconfigure into new project teams as a regular part of what we do—and working across boundaries is often a necessary condition of obtaining grant funding as we seek to make each proposal as strong and competitive as it can be. In addition, each of the team members had some experience—albeit in some cases minor—in the “other” methodology and therefore was technically able to participate meaningfully in their area of non-expertise. Particularly helpful in this regard was that one of our qualitative team members and one of our quantitative team members had some substantial experience in the other
methodology. These team members often served as “bridges,” providing explanation and connection across methods when necessary. Finally, through our grant we had access to methodological advisors—in both qualitative and quantitative methods, but importantly, in mixed methods through our work with a mixed methods specialist. While these advisors certainly provided advice to us about the particulars of our study, we also viewed our interaction with them as a professional development opportunity for all of us to learn together as a team. This was particularly important in the area of mixed methods, where we all had a common desire to learn more about the state of the art and to push our study to integrate methodologies purposefully and meaningfully.

**Publishing Findings**

The problem of how to sub-divide a large, multilevel project for publication reflects concerns addressed in the mixed methods literature about the publishing a mixed methods project (Stange, Crabtree, & Miller, 2006). Our dissemination plan benefits from our mixed methods design, as it increases the range of potential avenues for publishing our findings. We can imagine that the findings of this study will appeal to a diverse audience, including those who are interested in the content findings of the study, the methodological approaches, and the public policy implications. With this in mind, we have planned a set of potential journal articles that reflect these possible audiences, and designated a point person for that publication based on their expertise. For each publication, we also include team members who are interested in pursuing the results related to that journal and who can provide support for publishing that article. In addition, we have added a venue for discussion of lessons learned in the process of completing our large-scale, multilevel, mixed methods study that we did not initially plan to pursue. Table 3 is our
internal planning memo (subject to change, of course) that reflects this diversity of audiences and writers.

Challenges

In this section, we highlight four major challenges we encountered in implementing our multilevel design. These included creating and maintaining parallel constructs across both datasets; managing large, multilevel datasets; managing the differing sample sizes and data sources at our points of interface; and articulating and reporting on our methodology.

Creating and Maintaining Parallel Constructs

As discussed earlier, our understanding of the relation among our constructs of district-level support, school-level support, teacher use, and student outcomes came from previous qualitative research (Mark, Spencer, Zeringue, & Schwinden, 2010; Spencer, Mark, Zeringue, & Schwinden, 2010) and experience working with many school districts in curriculum selection and implementation. In this study, we sought to extend our understanding of these constructs through large-scale, comprehensive survey data as well as further qualitative exploration. Because of the origin of the constructs, they were relatively easy to operationalize on the qualitative side but much more difficult to operationalize on the quantitative side.

For example, we knew from our qualitative work that districts can make various “moves” to support teachers as they implement mathematics instructional materials, such as purchasing the materials needed, creating a pacing guide, offering professional development, and specifying the amount of instructional time on mathematics required each day. However, we also know the implementation of such strategies can vary at the school level—some school leaders are more effective than others at executing these moves, and the district itself may choose particular schools in which to deploy scarce resources (e.g., a mathematics coach or specialist) in a given
year. School leaders can also initiate additional supports using their own resources, particularly when support is lacking at the district level. Qualitatively, it made sense for us to analyze district- and school-level support separately—despite the overlap in the type of support provided at each level—because we could explain how those supports function differently at each level. But the overlap of these support dimensions would have caused a measurement problem due to issues of multicollinearity in our quantitative models (as we could have—from a theoretical point of view—reasonably assigned items to both the district-level and school-level support dimensions). We struggled to operationalize the constructs and ensure that each was operating independently. We decided to assign support variables that are more likely to vary by school such as teacher professional development to the school level and to assign variables that the district specifically controls such as presence of district pacing guide to district level.

Managing Large Multilevel Datasets

A significant challenge on this project has been managing the large amount of data required by our design and collected from 2,172 teachers on four teacher surveys, 153 principals on two principal surveys, 137 school and district staff through interviews, and 12 site visits. The challenge of managing large amounts of data is not unique to our study, but we found it particularly demanding to reduce and analyze the data in parallel in order to be able to merge databases and analyze convergence. Throughout the data cleaning and reduction phase, we carefully kept track of which variables codes were related to which constructs and dimensions so that we were able to identify comparable data. We also had to ensure that the aggregate variables or indexes we created with the quantitative data aligned and were comparable to the qualitative coding and memos. Similarly, we had to be sure not to reduce either dataset too much so that at the point of interface we lost the ability to fully merge the data or detect the level of convergence
(e.g., some data reduction for our quantitative dataset occurred after the two datasets were compared in our merged analysis).

Managing Different Sample Sizes and Sources at the Points of Interface

In the Points of Interface section above, we discussed the three ways we are integrating our qualitative and quantitative datasets. One of the challenges we faced in integration was managing the different sample sizes and multiple sources for our qualitative and quantitative data. At the district level, we have a large amount of qualitative data from each site that includes interviews with district leaders, site visit notes, and district documents. The district-level quantitative data includes items on teacher and principal surveys that relate to district-level support. While the constructs and dimensions that the data address are the same, the sources and the sample for the two datasets are different and represent different perspectives. Although this is a strength for our project—we believe the merged analysis of district-level support is stronger because various perspectives and sources create a more complete picture together—it was a practical challenge to manage the multiple sources and samples.

For school-level support and teacher use, we have a large amount of quantitative data (surveys from 2,173 teachers and 153 principals) and a relatively small amount of qualitative data (interviews with teachers or principals from 40 schools). At this level, the participants in the qualitative sample are also participants in the quantitative sample, so we can directly compare these datasets. For the explanatory analysis, we kept all 111 interviews from all 40 schools in the qualitative dataset, because we used that data to explore themes and relationships among our constructs; this explanatory purpose is not dependent on a minimum number of interviews per school. For the convergent analysis at the school level, we used the qualitative data to check the validity of our quantitative data at the school level and, therefore, only included the 23
qualitative subsample schools from which we have three or four interviews. In the convergent validity analysis, the qualitative sample is compared with the quantitative sample of teachers from each school (generally over 70% of the teachers in the school); for example, the qualitative data from three teachers is compared with the quantitative data from 12 teachers. The number of interviews per school matters much more in considering whether the data were comparable. As we did these comparisons, we were careful to keep in mind that the qualitative data reflect a smaller portion of the school staff and to consider to what extent their experience reflects the experience of all the teachers. Attending to the sources and sample sizes—and how they differed across the points of interface—was critical to integrating the datasets meaningfully.

Articulating and Reporting on Our Methodology

Although several of our staff members had some experience working across methodologies, this project led us all deeper into the field of mixed methods than we had ever traveled before. It was very helpful to us to have a guide on that journey; our methodological advisor helped us to see how our study related to the state of the art in the mixed methods field, and how it related to work that was happening outside our discipline. It was particularly helpful to have his guidance as we worked to articulate our multilevel design and clarify the purpose of our mixed methods approaches at each level. Writing this methods paper was an additional venue for articulating that design; it was a new experience for the research team, as most of us had previously focused on writing only in our disciplines. We found the experience valuable for several reasons. First, it helped us to clarify the methods and strategies we used in all phases of our research and to reflect on our purposes in choosing these strategies. Second, undertaking the work of writing this paper has expanded our knowledge and understanding of the mixed methods field—in particular, of the technological innovations that facilitate comparisons across data types
and the work that is done in mixed methods outside of the field of mathematics education.
Finally, we are pleased to be able to contribute to the understanding of how mixed methods can be applied to large multilevel studies.

**Significance**

This project contributes to the knowledge in both mixed methods research and in mathematics education. The contribution to the mixed methods literature is to advance the design of a multilevel mixed methods project in which data were gathered at several levels, merged, and analyzed to assess the level of convergence and to explain themes and relationships among constructs. It also advances challenges likely to arise in this form of complicated design: early decisions about teamwork and the consistent use of constructs and parallel questions for both the quantitative and qualitative arms of the study, the specific types of data and where and how they might be used in tandem in the project, and how to analyze the data so that one can augment the other and so that contradictions between the data sources can be addressed. It also explicates the procedures we used so that others might have some guidance in conducting multi-level designs.

In mathematics education, this study adds to the literature by investigating and operationalizing several key dimensions that constitute a coherent implementation strategy from a district perspective, considering the mediating role of school-based supports and level of use of the materials, and examining the connection between coherent implementation and student outcomes as measured on state assessments. It also validates constructs of school-level support and teacher use of materials. The results will assist district and school administrators, as well as policymakers at various levels, in planning implementation of mathematics instructional materials. This project contributes to the work of STEM education researchers by clarifying the
dimensions of coherent implementation of instructional materials and offering ways to operationalize and measure these dimensions.
References


Figure 1. Relation Among the Constructs

District-level Support
12 districts, 5 states

School-level Support
153 schools

School-level Use
2172 teachers

School-level Student Outcomes
fourth-grade state test
5 states

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.
Figure 2. The display represents our four key constructs, our quantitative sources of data for each construct, our qualitative sources of data for each construct, and the points of interface between the two datasets for each construct.
Figure 3. Resolving Discrepancies

Figure 3. This figure shows the steps we took in this analysis of our school-level joint displays.
Table 1

*Use of Parallel Questions for Constructs: Pacing Guide*

<table>
<thead>
<tr>
<th>QUAN Survey Questions</th>
<th>QUAL Interview Questions</th>
</tr>
</thead>
</table>
| Does your district provide mathematics pacing guides and/or instructional charts that offer guidelines about which *Everyday Mathematics* unit to teach and when to teach it? | Is there a pacing guide in your district?  
Probe:  
- Do you follow it? Are you able to follow it as you get close to the testing window?  
- Does the principal or other person monitor your progress as compared to the pacing guide? Are you offered support if you need it?  
- Does the pacing guide follow the published materials or does the pacing guide pick and choose lessons/units.  
- Does the pacing guide include lessons or activities from other materials? If so, what? How frequently does this happen?  
- Is that helpful to you? Would you rather follow the book/pick and choose units? |
| - Yes                                                                                   |                                                                                           |
| - No                                                                                   |                                                                                           |
| In general, are you able to keep pace with the instructional chart/pacing guide that your district provides for *Everyday Mathematics*?  
- Yes.  
- No, the pacing guide is too fast.  
- No, the pacing guide is too slow.  
- Our district does not have an instructional chart/pacing guide. |                                                                                           |
| How do you use the mathematics pacing guides and/or instructional charts provided by the district?  
- I follow them very carefully, and I rarely deviate.  
- I follow them as best I can, but I occasionally deviate.  
- I use them as a guideline, but I sometimes deviate.  
- I look at them occasionally, but I regularly deviate.  
- I know the district has them, but I never use them.  
- Our district does not provide pacing guides or instructional charts. |                                                                                           |
### Example of Joint Display of Quantitative and Qualitative Data for a School

<table>
<thead>
<tr>
<th>School Support</th>
<th>QUAN: Surveys</th>
<th>QUAL: Interviews</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Necessities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Pacing Guides</td>
<td>a. All teachers ((N=12)) reported that they have a pacing guide.</td>
<td>a. There is a pacing guide.</td>
<td>a. QUAL and QUAN confirm that there is a pacing guide.</td>
</tr>
<tr>
<td></td>
<td>b. 75% ((N=9)) reported that they can keep pace, and 25% ((N=3)) responded that the pacing guide is too fast. All first- and second-grade teachers and one of the three third-grade teachers responded that they are able to keep pace. Two other third-grade teachers and the fourth-grade teacher responded that the pacing guide is too fast.</td>
<td>b. Curriculum specialist (CS) reported following the pacing guide has been challenging, but the teachers reported that they are not behind much.</td>
<td>b. While the curriculum specialist reported that it is challenging to follow the pacing guide, 75% of teachers responded they can keep pace. Based on QUAN data, the teachers in grades 1 and 2 are on pace while some of the upper-grades teachers thought the pacing guide is too fast.</td>
</tr>
<tr>
<td></td>
<td>c. 8.3% ((N=1)) reported that they follow the pacing guide very carefully and rarely deviate; 58% ((n=7)) reported that they follow the pacing guide as best they can but they occasionally deviate; 33.3% ((n=4)) reported that they use it as a guideline but sometimes deviate.</td>
<td>c. Pacing guide was followed strictly in Years 1 and 2. In Year 3, the new superintendent allowed more supplementation but there has been confusion about what and when to supplement and who decides. District says they told teachers they could supplement in October, but during site visit schools reported they didn’t hear that message until February.</td>
<td>c. QUAL and QUAN indicate teachers do supplement.</td>
</tr>
</tbody>
</table>
Table 3

Example of Collaborative Investment in Publishing from the Project

<table>
<thead>
<tr>
<th>Title/Focus of Paper</th>
<th>Target Journal</th>
<th>Data Sources</th>
<th>Who’s Responsible?</th>
<th>Deadline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coherent Implementation of Mathematics Instructional Materials: A multilevel mixed methods study</td>
<td>A methodology journal in mixed methods</td>
<td>Discussions of methodological occurrences Joint displays and analysis</td>
<td>DS, JC, KR, JY This includes a PI, our mixed methods advisor, a qual researcher, and a quant researcher</td>
<td>April 2013</td>
</tr>
<tr>
<td>Coherent Support for Elementary Mathematics Materials: Impacts on teacher use and student outcomes</td>
<td>A mathematics education research journal</td>
<td>Joint displays Quantitative analysis</td>
<td>DS, JM, KR, JY The two lead authors are the PIs; team also includes a qual researcher and a quant researcher</td>
<td>July 2013</td>
</tr>
<tr>
<td>Impact of District Policy on Elementary Students Mathematics Achievement</td>
<td>An education policy journal</td>
<td>Substudy on district coherence Quantitative analysis of student outcomes</td>
<td>KR, JY, DS The lead author is a qual researcher supported by a quant researcher and a PI</td>
<td>July 2013</td>
</tr>
<tr>
<td>Using Interrupted Time Series to Investigate the Effects of Support for Curriculum Implementation on Student Achievement</td>
<td>An education research journal focused on empirical studies</td>
<td>Quantitative analysis highlighting use of interrupted time series methodology</td>
<td>JY, JG, KR, DS The two lead authors are quant researchers, plus a qual researcher and a PI</td>
<td>August 2013</td>
</tr>
<tr>
<td>Teacher Use of Mathematics Instructional Materials: Patterns across grade-level and time of year</td>
<td>A professional journal for practitioners in mathematics leadership positions</td>
<td>Joint displays Descriptive analysis of use quantitative findings</td>
<td>DS, KS, KR, JM, JY The lead author is a PI, plus the qual researchers with support from the quant researcher</td>
<td>TBD</td>
</tr>
</tbody>
</table>