Improving Mathematics Teacher Practice and Student Learning through Professional Development

PROJECT DESCRIPTION

We propose to study how to provide, on a large scale, the professional development and continuing support teachers need to improve student achievement in mathematics. Over five years of a longitudinal study, we will examine how the use of specific research-based instructional strategies in the classroom—supported by professional development and highly rated curriculum materials—relate to lasting improvements in student learning. An experimental study in Years 4 and 5 will test the feasibility of delivering professional development and ongoing support cost-effectively on a large scale. The University of Delaware and Texas A&M University, working in partnership with Project 2061 of the American Association for the Advancement of Science (AAAS), will address Focus Area II of the IERI program solicitation: Transition to Increasingly Complex Science and Mathematics Learning. The research questions for this proposal are designed to study the interactions of teaching practices, curriculum materials, and professional development to understand how they can be optimized to improve student learning. (See letters of support in Appendix A.)

Theoretical Foundations and Rationale

Improving student learning is an extraordinarily complex problem, requiring coordinated reform of many different parts of the educational system. Such reform involves setting coherent learning goals; creating aligned assessment; developing sound curriculum materials; dramatically improving the quality of teaching; and allocating adequate conceptual, material, and social resources in school settings (Fuhrman, 1993; Fullan, 1991; AAAS, 1998).

A large body of research has identified elements of effective instruction in mathematics including eliciting and responding to student conceptions, structuring inquiry with a series of mathematically-rich problem situations, and promoting discussions among all students in which correctness lies in mathematical argument rather than the social status of the participants (e.g., NRC, 2000; AAAS, 1989; Grouws, 1992; Hiebert et al., 1997). However, we do not know what sense the majority of teachers make of these elements and some case studies report that teachers use the rhetorical language of reform but never actually incorporate new and more effective instructional practices (Cohen et al., 1990). And there are almost no empirical studies on how to develop on a large scale teachers’ knowledge and skills needed to enact effective instructional practices.

It has been argued that curriculum materials could be agents of instructional change and sources of teacher knowledge (Ball & Cohen, 1996). Indeed, Project 2061’s recent evaluation of middle grades mathematics materials indicates that the newly developed materials funded by the National Science Foundation have high potential for improving student learning (AAAS, 2000). Yet while a few studies suggest that use of these materials improves student achievement, most studies were carried out by the developers and lacked control groups (Show-Me Center, 2000). Furthermore, these studies did not examine how typical teachers used these materials in real classrooms and in what respects the materials contributed to the improvement of teachers’ knowledge. Given the potential importance of curriculum materials (Ben-Peretz, 1990; Freeman & Porter, 1989; Stodolsky, 1988), it is fundamental to understand how they can support teachers in ways that allow them to actually understand and implement effective teaching practices.

School districts adopting new research-based materials are often unaware of the magnitude of change in teachers’ thinking and practice that is needed to implement them effectively. Enthusiasm for these materials often wanes when implementation difficulties ensue. Without professional development explicitly aimed at helping teachers both understand the intent of these curriculum materials and assess models of what it looks like to use them well, it is unlikely that school districts will be able to exploit the full potential of these
materials for increasing teacher knowledge and student learning. But professional development by itself is unlikely to produce change, in the same way that curriculum materials alone or teachers alone are unlikely to improve student learning. We contend that it is the interactions among teachers, curriculum materials, professional development, and ongoing support for teachers that can lead to lasting improvements in student learning. To examine these interactions, the proposed project will focus on three questions:

1. **What is the relationship between teacher knowledge, the use of research-based instructional strategies—supported by highly rated curriculum materials—and student learning of specific mathematics ideas and skills?**

2. **How does professional development and ongoing support—focused on specific mathematics learning goals—build teacher knowledge and lead to more effective teaching practices?**

3. **How can technology help to provide effective teacher professional development and ongoing support cost-effectively on a large scale?**

The starting point for addressing these questions is a set of research findings and assumptions that relate to what students should learn, how teachers and students gain new knowledge and skills, how curriculum materials support good teaching, and how specific kinds of professional development and ongoing support can change teacher practices and sustain those changes over time.

**Mathematics knowledge and skills.** There is now widespread agreement that mathematics education should focus on a carefully specified and coherent set of important concepts and skills that all students should learn, such as the learning goals (or standards) recommended in the National Council of Teachers of Mathematics’ (NCTM’s) *Principles and Standards for School Mathematics* (2000) and in Project 2061’s *Benchmarks for Science Literacy* (AAAS, 1993). These documents are more than a list of objectives; rather, they provide a coherent picture of what all students should know and be able to do at the end of their K-12 education. They are a fundamental framework for education research programs. When learning goals are spelled out with sufficient specificity and coherence (as these national goals are), it becomes possible to systematically study how students think about particular mathematical ideas and skills and what kinds of teaching approaches are effective in helping students learn them. Indeed, much of the knowledge needed for teaching is specific to a particular idea or skill (Shulman, 1986; Carpenter et al., 1989).

**Teacher knowledge.** Research has illuminated the kinds of teacher knowledge needed to teach specific mathematical ideas and skills for understanding (e.g., Ball, 1988; Leinhardt & Smith, 1985; Putman & Borko, 2000; Fennema & Franke, 1992; Cohen et al., 1993; Hiebert & Carpenter, 1992). According to a forthcoming report from the National Research Council, mathematics teachers need specialized knowledge that “includes an integrated knowledge of mathematics, knowledge of the development of students’ mathematical understanding, and a repertoire of pedagogical practices that takes into account the mathematics being taught and the students learning it” (NRC, 2001a). Teachers also need content that is much more coherent than the traditional content portrayed in widely used textbooks. In addition, students need a collection of representations to help them construct their own mathematical understanding. Content-based pedagogy helps teachers to recognize common student conceptions that are relevant to specific learning goals and to provide tasks and pose questions that guide students’ interpretation of mathematics (NRC, 2001a). These and other principles are embedded in the *How People Learn* report (NRC, 2000) and Project 2061’s research-based criteria (see Appendix B) for evaluating key elements of effective mathematics instruction (AAAS, 2000). The criteria require materials and teachers to provide a sense of purpose for students, build on students’ ideas, engage students in mathematics through a variety of contexts and experiences, develop and apply mathematical ideas, promote students’ thinking about mathematics, assess student progress, and enhance the learning environment for ALL students. These criteria will be used in our study to (1) inform observations of teacher practices and relate these practices to teacher knowledge and (2) shape the design of professional development and ongoing support for teachers.
**Student learning.** In creating standards documents, both NCTM and AAAS drew on a large body of research findings on how students learn particular ideas, when they are best able to learn them, and the difficulties they are likely to have (e.g., Owens & Wagner, 1993; Hiebert & Carpenter, 1992; Carpenter et al., 1989; Cobb et al., 1991). Among the most significant findings are those that deal with how students (indeed, how all learners) connect new information and concepts to those they already have (NRC, 2000). Current research has much to say about students’ common preconceptions and how students restructure their thinking to incorporate new ideas and skills, how they construct new knowledge, and how they transfer knowledge to new situations and applications. These findings have also shaped Project 2061’s research-based criteria for evaluating instructional strategies in the classroom and in curriculum materials (AAAS, 2000).

**Curriculum materials.** The mathematics concepts and skills described in the NCTM standards and Project 2061’s benchmarks have helped to guide the development of new textbooks, including several that have been created with support from the National Science Foundation. To judge their potential in helping students learn concepts and skills included in standards and benchmarks, Project 2061 conducted an independent evaluation of 13 recently published middle grades mathematics textbooks. Drawing on the available research on how students learn mathematical concepts and skills, Project 2061 crafted a set of criteria for judging the quality of the instructional support provided in curriculum materials (AAAS, 2000). While nearly all of the textbooks included in the evaluation provided adequate coverage of selected mathematics ideas and skills, only four texts were rated satisfactory in providing effective instructional strategies for teachers, receiving an overall score of at least 2.0 out of a possible 3.0 (AAAS, 2000). These books are carefully aligned with national standards and draw on learning research to provide effective instructional guidance that supports (but does not replace) the teacher’s own knowledge of the mathematics content and how to teach it. Our research assumes that curriculum materials affect student learning mainly through their influence on teachers. Rather than trying to “teacher-proof” materials, we assume that highly rated materials support teachers by helping them to build their own content and pedagogical knowledge and to enact and reflect this knowledge in their teaching. Highly rated materials provide a powerful set of carefully selected and sequenced tasks to evoke student conceptualization and reconceptualization of important mathematical ideas. These resources can serve as sources of reflection for teacher learning as well (Smith, 2001; Acquarelli & Mumme, 1996; Ball & Cohen 1996).

**Professional development.** There is a growing consensus among educators on the qualities and goals of effective professional development (Loucks-Horsley, Hewson, Love, & Stiles, 1998; NRC, 2001a, 2001b; Wilson & Berne, 1999; Ball & Cohen, 1999; Smith, 2001). For example, recent studies indicate that successful professional development programs in mathematics are content-driven and explicitly focused on how students learn particular mathematics ideas (Kennedy, 1999; Smith, 2001). Also, there are reports that support the principle that professional development should be relevant to teachers’ practices, long-term, content- and community-focused, and collaborative (Wilson & Berne, 1999; NRC, 2001a, 2001b). Research has shown the importance of inquiry as part of teaching practice (e.g., teachers collecting and analyzing data on their own students’ learning) and the value of ongoing support for teachers through mentoring and through peer group planning, discussions, and lesson studies (Carpenter et al., 1989; Kennedy, 1999; Wilson & Berne, 1999; Silver & Stein, 1996; Smith, 2001). Unfortunately, there is almost no research on how professional development actually affects student learning (Kennedy, 1999).

As the findings summarized above show, researchers have produced some knowledge about how different elements of the educational system—such as curriculum, instruction, professional development, and ongoing...
support for teachers—should function to improve student learning of mathematics. But there is no research on how all these elements should work together (Cohen & Ball, 1999; Spillane, Halverson, & Diamond, 2001). By examining the connections among the structure of instructional materials, teacher knowledge, classroom activities, professional development, ongoing support, and student learning, our research will shed light and provide valid statistical evidence on how these elements work together to improve student learning in mathematics. We have designed and plan to execute a rigorous longitudinal study. We will create and adapt research-based instruments to collect valid teacher and student learning data that can provide information on the conditions under which students learn mathematics well. In particular, we propose to study the classroom conditions that enable students to achieve the ambitious learning goals set forth in the new generation of reform curricula. Ultimately, these findings will strengthen national policy decisions about the role of curriculum materials development, professional development, and ongoing support in promoting student achievement in mathematics. The project takes advantage of the variety of development and implementation efforts that currently exist in mathematics education and addresses key questions asked by educators and the public: Can the reform curricula really improve student learning? Under what conditions does such learning occur?

Research Design
The research design of the project is inevitably a compromise between research logic and the practicalities of the sites involved. Perturbation in the context at any one site may require some adjustment in the details of the design. To evaluate the effectiveness of our professional development interventions, we will (1) conduct classroom observations to document teacher and student behaviors in the classroom as they work with materials that have received high, moderate, and low ratings and (2) assess student learning of key concepts and skills that are widely regarded as important to basic understanding in mathematics. We will work with schools in two states where standards-based reform efforts have been underway during the past decade. Both states have content standards that correspond well to the learning goals that will serve as the basis for this project.

Research Questions in Detail
By pursuing the following three research questions, we propose to investigate the kinds of evidence (e.g., summaries of research, examples of teaching practice, student achievement results) teachers need to enable them to change their instructional practices in order to teach for understanding:

1. **What is the relationship between teacher knowledge, the use of research-based instructional strategies—supported by highly rated curriculum materials—and student learning of specific mathematics ideas and skills?** Materials earned high ratings on the Project 2061 instructional criteria if they provided explicit and high quality support for effective teaching toward the learning goals examined (AAAS, 2000). Our goal is to investigate what these highly rated materials contribute to teacher knowledge, how they influence teacher practice, and how teachers’ use of the support provided in the materials relates to student learning. Describing and explaining the relationships among teaching practices, the use of curriculum materials, and student learning is a complex research challenge requiring the collection of data for different indicators of each of the relevant variables. In the Research Methodology section below, we describe the indicators of teacher knowledge and behavior and of student learning that will be used as a basis for these investigations.

2. **How does professional development and ongoing support for teachers—focused on specific mathematics learning goals—build teacher knowledge and lead to more effective teaching practices?** Professional development and ongoing support will explicitly focus on the learning goals, alignment of curricular content to them, research-based teaching toward them, how highly rated materials support teaching them, and student learning of them. We will collect teacher and student data as evidence of the impact of this kind of goals-based professional development on teachers’ instructional practices. We plan to explore this question developmentally during a five-year longitudinal study, documenting how
teachers’ knowledge evolves during the first three years of professional development and subsequent years of ongoing support and how this relates to the implementation of research-based teaching practices. We will design the professional development and ongoing support to respond to observed deficiencies in knowledge and practice.

3. How can technology help to provide effective teacher professional development and ongoing support cost-effectively on a large scale? Face-to-face professional development and ongoing support, while having many advantages, is labor intensive, difficult to schedule, and impractical to deliver on a large scale. For example, a district implementing a research-based teaching unit supported novices by placing an experienced user (a pilot test teacher) in each school (Lee et al., 1993). But this is not feasible for most schools implementing new materials. Technology can provide a partial solution by spanning distance and time. We will examine the ability of telecommunications technology such as the Web to support conversations between novice and more experienced teachers in widely separated schools or districts and facilitate their analysis and discussion of sample lessons. In addition, we will explore how technology can be used to support transfer of knowledge and skills to new learning goals (e.g., by supporting the exchange of video clips of lessons and samples of student work resulting from those lessons) and sustain a district’s use of research-based teaching practices (e.g., by supporting induction of new teachers into communities of practice). We will employ LessonLab software and consultants to facilitate lesson analysis.

Case-Based Interventions for Professional Development and Ongoing Support

In keeping with our view of teaching as scientific inquiry, we think improving classroom practices is best achieved and sustained by helping teachers learn how to collect and use data—about their practices, about their students’ learning, and about how these two elements interact. At the center of our interventions for professional development and ongoing support are case studies drawn from the data that researchers and teachers will collect. The use of case studies can encourage flexible adaptation to teachers’ own students and contexts and accelerate the normally quite lengthy process of experience acquisition (Spiro & Jehng, 1990; Spiro, Feltovich, Jacobson, & Coulson, 1995). Case studies have been successfully used in mathematics education (Shulman, 1992; Lampert & Ball, 1998; Stigler & Hiebert, 1999). Our case studies will be designed to help teachers improve their classroom practices so that they meet research-based criteria for effective teaching toward specific learning goals and the principles from which they are derived. Case studies will include samples of teaching focused on specific learning goals (e.g., lessons presented in the curriculum materials involved in this study, videos of enacted lessons, transcripts of enacted lessons or parts of them) and resulting student learning (e.g., samples of students’ work on relevant problems, transcripts of interviews with students, scores on test items). Focused discussions of case studies during professional development will clarify the kinds of thinking and practices desired—in planning, carrying out, and assessing lessons (Smith, 2001). Similar discussions will be encouraged and supported at school sites and online.

The professional development and ongoing support are based on and will model the same research-based principles of effective teaching that we want teachers to include in their practice. Specifically, the proposed professional development will do the following:

Provide a sense of purpose. Rather than presenting principles of effective teaching as abstractions, professional development will be situated in relevant classroom teaching situations (Wilson & Berne, 1999; Ball & Cohen, 1999; Smith, 2001). For example, student learning goals will be examined in the context of selecting lessons that align with them and in considering what student achievement of those goals would look like. The purposes of professional development activities will be explicitly stated and strategically sequenced to accomplish these purposes.

Take account of initial ideas and skills. Professional development will attend to prerequisite knowledge
and skills, taking time to find out where participants are starting from and helping them overcome both common misconceptions about mathematics learning and ineffective habits of teaching. For example, some teachers may view mathematics as little more than rules and procedures and mathematics teaching as helping students learn to follow them rigidly (NRC, 2001a; Stigler & Hiebert, 1999). Professional development will make use of student interviews and samples of student work, helping teachers to appreciate what each reveals about students’ underlying conceptual understandings and what the negative consequences might be of insisting students with those understandings follow rigid procedures. We will use the mathematics maps from Atlas of Science Literacy (AAAS, 2001) to help teachers understand what prerequisites are important to take into account in teaching an idea or set of ideas or skills. (See Appendix C for an example of a relevant map.)

**Provide experiences with relevant phenomena.** By phenomena, we mean observable events that can help make the ideas or skills to be learned credible. For professional development, the events are instances (identified in case studies) of proficient teaching toward specific learning goals or instances where proficient teaching correlates well with improved student achievement of those goals. As teachers collect and analyze similar kinds of data in their own classrooms, their observations will help validate their use of research-based instructional strategies (Smith, 2001).

**Develop and use ideas and skills.** Professional development will use data collected to develop an evidence-based argument for the merit of ideas and skills to be learned, model the use of research-based teaching practices (in case studies), discuss essential aspects of the models and criteria by which teachers can analyze their own teaching, and provide opportunities for teachers to practice skills in their own classrooms and obtain feedback (e.g., by bringing in and discussing video excerpts of their own classroom practices).

**Promote thinking about experiences with phenomena, knowledge, and skills.** Professional development will provide opportunities for teachers to explain their ideas and demonstrate their skills to one another (such as with video excerpts of their own teaching) (Franke, Carpenter, Levi, & Fennema, 1998); guide their interpretation and reasoning about mathematics ideas, learning, and teaching (e.g., helping them to see beyond details of instances to general principles of learning and teaching); and encourage them to monitor their own progress (such as by helping them to compare their initial and subsequent attempts to employ research-based strategies). Teachers will work in groups and will be encouraged to continue these focused group discussions offsite, both face-to-face and electronically.

**Assess progress.** Teachers’ learning will be monitored by questionnaires and interviews, observations of their classroom teaching practices, and analyses of their students’ learning. In this context, teacher learning includes understanding the targeted content and pedagogical skills, being able to apply them in familiar and novel situations, and being inclined to do so. Findings will inform revisions in the design of professional development and ongoing support.

**Enhance the learning environment.** Professional development will help teachers improve their own understanding of mathematics ideas, learning, and teaching. For example, discussions about students’ ideas in the context of samples of their work can help teachers recognize some of their own conceptual deficiencies and deal with them. Sessions will create a classroom environment that welcomes teacher curiosity, encourages a spirit of healthy questioning, avoids dogmatism, encourages high expectations, and enables every participant to experience success. In this way, professional development will “create contexts for teacher collaboration, provide a focus for the collaboration, and provide a common frame for interacting with other teachers around common problems” (NRC, 2001a, pp. 10-28).

Just as we want to encourage teachers to use feedback from student learning to inform decisions about curriculum and teaching, so our professional development and ongoing support will be revised—during the longitudinal study and in preparation for the experimental study—based on feedback on teacher
learning. We will use teacher feedback, what can be inferred from classroom observations, and the availability of videos of teaching and samples of student work from the first three years to design and test an optimal professional development experience, supported with technology, for half of the new participants in Years 4 and 5. In the interest of scalability, we will also design and test a highly condensed (minimal) professional development experience that can be delivered in a shorter time frame, but which is supported by the same technology.

**Focus of professional development and ongoing support.** Initially, professional development will focus on the learning goals themselves and what “alignment” to them entails. Teachers will identify relevant activities in the four curriculum materials involved in this study (see Methodology section for a description of the materials) comparing how well the activities focus on the learning goals and how the activities are distributed over time. (See Appendix D for a list of professional development objectives and activities.)

With the learning goals well in mind, professional development will focus on developing an understanding of research-based instructional practices, how highly rated materials support teachers in applying these practices in the context of specific learning goals, and how these practices relate to student achievement. Participants will examine artifacts of teaching—for example, analyzed lessons (as characterized in the curriculum materials)—and compare how well they reflect particular research-based principles (Stigler & Hiebert, 1999). As we collect a range of enactments of these same lessons from participating teachers’ classrooms, we will involve teachers in analyzing the enacted lessons. We think that the range of both analyzed lessons and analyzed enactments will contribute to developing the cognitive flexibility teachers need to recognize the essential attributes of good instruction and the varying ways those attributes are combined and adaptively tailored to diverse classroom situations and then to examine their own practice in light of them (Spiro, Coulson, Feltovich, & Anderson, 1994). We will attempt to balance the use of live videos (to capture a wider range of behavior and provide greater interest) and the use of transcripts of those videos (to allow for greater focus and preserve anonymity). As the relationships emerge from the studies between the use of research-based practices and student achievement of the specific learning goals that these practices are aimed at, we will involve teachers in examining and discussing the data. For example, student work on a particular task in “highly rated” classrooms (where the enactment meets particular instructional criteria) will be compared to student work in less highly rated classrooms (where, for example, teachers pay less attention to commonly held student ideas or to guiding student interpretation and reasoning about the task).

**Development of case studies.** Some case studies have already been developed and tried out as part of Project 2061’s ongoing professional development programs (AAAS, 1997; Roseman, 1997; Kulm, Bush, & Surati, 2001). But most case studies will be developed to illustrate proficient teaching toward standards and how to analyze lessons and student work to improve teaching and learning. For example, we will use videotaped lessons of classrooms using highly rated mathematics materials (and corresponding transcripts of those lessons using LessonLab software) and help teachers analyze how the lessons support learning of important ideas and skills. LessonLab software will make it possible for professional development leaders and participants to study videos and transcripts of the same lessons and to annotate them for subsequent discussions.

**Ongoing support.** Given the limited time available for professional development and the relatively large number of learning goals to be achieved in middle grades mathematics, it is not possible to provide case studies that address each student learning goal. We hypothesize the highly rated mathematics materials can help support teachers’ attempts to apply research-based practices to other learning goals—something the proposed studies will explicitly test. But teachers will likely also need ongoing support to
understand the other learning goals and to adapt the materials in diverse settings. Furthermore, as new teachers enter the system they will need to become part of the communities of practice: “When teachers have opportunities to continue to participate in communities of practice that support their inquiry, instructional practices that foster the development of mathematical proficiency can be more easily sustained” (NRC, 2001a, pp. 10-28). We will encourage such communities of practice by
• establishing the value of such communities during professional development sessions;
• helping policy makers and parents appreciate the value of such communities to improving teaching and learning;
• helping all of them consider strategies for making time available for the analysis of practice and student learning to continue (e.g., Smith, 2001);
• annotating and making the case study materials (from this project) available for districts to use with other teachers (especially those new to the district);
• providing telecommunications networks to support online discussions during and beyond the project; and
• encouraging the use of findings and materials from this project in graduate courses.

A key aspect of sustaining improvements in teaching is the need to build district capacity for providing feedback and support (Spillane et al., 2001). We will include district mathematics specialists in participating and helping to plan professional development, collecting data, and analyzing data and make the case studies developed in this project available to participating districts (see letters of support in Appendix A). Telecommunications links will be used to disseminate case study materials and to connect mentor teachers with less experienced teachers and study groups with each other. We will also encourage participating universities to offer teachers graduate credit for their work on this project.

**Research Methodology**

Years 1-3 begin a longitudinal study of progressive professional development of a small number of teachers. To study persistence and sustainability, follow-up data are collected in Years 4 and 5. To test the adequacy of 20 days of condensed professional development, an experimental study begins in Years 4-5. A new cohort of teachers will be randomly assigned to receive either an optimal or a minimal version of the professional development, though both groups will receive technology-based support. The experimental study will evaluate how the professional development strategies employed so far can be abbreviated and scaled up to affect implementation and achievement more widely and more efficiently. We will measure the level of implementation and student achievement in the new teachers’ classrooms, looking again at how teachers improve implementation, how they adapt to individual needs, and whether teachers are able to compensate for shortcomings in lower-rated materials. The Year 4 students will be assessed in Year 5. (At some sites, it may be possible to track the achievement of Year 5 students in subsequent years through state test scores.) (See Appendix E for a flow chart of the studies.)

**Selection of learning goals.** We will choose learning goals from the conceptual strands of number, geometry, algebra, and statistics commonly found in grade 6-8 textbooks and will include both conceptual and procedural ideas. The learning goals (see Appendix F) will be selected in cooperation with the participating schools and teachers and will focus on: using, interpreting, and comparing numbers in equivalent forms, such as integers, fractions, decimals, and percents; identifying, comparing, and analyzing simple plane figures and solids and comparing shapes in terms of concepts such as parallel and perpendicular, congruence and similarity, and symmetry; using graphs to represent a variety of possible relationships between two variables; using symbolic equations to summarize how the quantity of something changes over time or in response to other changes; and using the graphic display of numbers to organize, analyze, and make predictions about data.

**About the textbooks.** Two of the textbooks—Connected Mathematics and Mathematics in Context—
were the most highly rated of the 13 analyzed by AAAS, with overall median scores of 2.7 and 2.8 out of a possible total of 3.0. One series—*Middle Grades Math Thematics*—received an overall median rating of 2.3, which was considered to be just above satisfactory. Moreover, while receiving an overall rating of satisfactory, *Middle Grades Math Thematics* received unsatisfactory ratings on three out of four criteria in Category II: Building on Student Ideas about Mathematics. The fourth series—*Mathematics Applications and Connections*—is widely used in Texas and other states but received an overall median rating of 1.35, falling in the unsatisfactory range. (See Appendix G for comparison of ratings of the four textbooks.) *Middle Grades Mathematics Textbooks: A Benchmarks-Based Evaluation* (AAAS, 2000) describes the analysis and provides the data. By using textbooks with this range of support for effective teaching, we hope to learn (from Year 1 of the study) what level of support is actually needed and (from Years 2 through 5) how much professional development can compensate for the lack of support provided.

**About the study sites.** The seven school districts involved in this study serve diverse student populations, ranging from 60% “Low SES” and minority in Bryan, TX to 10% each in Appoquinimink, DE. Typically, sites have about a third “Low SES” and 40% minority students.

Sites using *Mathematics in Context* and *Mathematics Applications and Connections* provide the most mature implementation examples. Teachers have at least three years of experience using *Mathematics in Context*. *Mathematics Applications and Connections*, while only adopted in 1999, is comparable to books previously used in the district. In contrast, *Connected Mathematics* and *Middle Grades Math Thematics* teachers have one or two years of experience with the series.

In Delaware, we will work with their Local Systemic Change initiative, which is already supporting the implementation of two of the highly rated mathematics materials, *Connected Mathematics* and *Mathematics in Context*. Delaware teachers using these materials will all have received 80 hours of professional development in their use. Texas teachers have had no professional development. By using two sites with different levels of prior professional development for the same material (*Connected Mathematics*) we hope to gain insights about the extent of this kind of professional development that is actually needed.

One of the sites in Texas is beginning to implement *Connected Mathematics* at the 6th grade level in one or two schools. The remaining schools in the district continue to use a low-rated material but are considering changing to *Connected Mathematics*, providing us with an opportunity to study the factors involved in making decisions about curriculum change. The other Texas district has been using *Middle Grades Math Thematics*, a moderately rated material, and has been providing teachers with short workshops in which they do sample activities from the textbook.

**Assignment.** Ideally, professional development would be assigned randomly to different teachers in subgroups cross-stratified by relevant external factors such as experience and subject knowledge. In the longitudinal study, random assignment will be made to volunteers (who are willing to be randomly assigned). In Years 4 and 5 we plan to use “waiting list” assignment, in which a random half of the teachers willing to participate are promised the same training at a later time. Half of the new teachers will be assigned at random to participate in optimal (up to 20 days) or minimal (one or two days) levels of professional development distilled from the professional development experiences of the earlier years. All teachers will receive technological support in the form of annotated video clips from early classroom observations, discussion guides, and telecommunication links.

Factors that cannot be controlled by assignment can be entered as covariates in the statistical analysis of results, although covariates may have distinctly non-linear effects and interact with one another. Where
randomization is not possible in the proposed study, teachers will serve as their own controls, with current students’ performance being compared to that of previous students. That would account for most of the contextual variables listed above. Recent advances in growth modeling such as latent growth modeling (largely an SEM-based method), hierarchical growth modeling (HLM-based), and longitudinal modeling (SEM-based) allow a much richer representation of change in teachers and students than has been typically been reported (McArdle & Hamagani, 1997; Raudenbusch, 1988; Sivo & Willson, 2000; Willett & Sayer, 1997).

Analysis. Years 1-3 of instruction begin a five-year longitudinal study of the same set of 50 teachers, in which their effectiveness is studied over three successive years of professional development and practice. The entering characteristics of the three successive student cohorts can be recorded over the three years, providing an estimate of the year-to-year variance in student population. Growth of student understanding will be assessed at three time points in their year in the class—the first an early pre-test, the middle immediately after instruction in the unit being studied, and the last near the end of the school year. This last will be a test of further growth, or a medium-delay test of retention. Where school policy allows, students in the first year can be assessed for retention at the beginning of each of the following two years (providing further points on the learning curves). These multiple time points will provide not just a summary of student learning, but estimators of the shape of the learning curve. Alternative forms of assessments will be needed.

Student understanding accumulates over time and also undergoes some predictable restructuring. Such knowledge restructuring will be evaluated through careful selection and development of assessment tasks to uncover patterns of student understanding (such as progress along conceptual maps, see Appendix C). Latent growth models using SEM and HLM will provide sufficiently sophisticated methodologies to encompass the effects of instruction (teacher variation and change), student knowledge structure (initial conditions), and student change in knowledge structure (amount and form). These methodologies are now well established and functional using commercially available statistical packages. With at least 10 instructors using each material and 200-300 students per year, statistical power will be adequate to detect small instructional effects. After two years, estimation of instructional effects will allow confirmatory analyses for Year 3. This methodology has not been commonly conducted in educational research, yet it is parallel to methods in science using prediction to support validity of effects.

Classroom data collection. The primary purpose of classroom data collection is to describe actual instruction (enacted curriculum) so that it can be held up to the same principles of effective teaching that were used to analyze the support provided in curriculum materials. A detailed description of the program of activities and resources recommended in each material—that is, the literal program—provides a basis for monitoring, describing, and analyzing teachers’ use of the material, noting, for example, which tasks are actually implemented in the classroom, what modifications are made, and what additional tasks are assigned. (See Appendix H for an example of the instrument.)

The monitoring of teachers’ use of the materials will be done in a way that respects the teachers’ responsibility to adapt the materials to a particular group of students. This will shed light on the kinds of modifications that teachers judge are needed. Our interest is not in making sure that teachers follow the material rigidly; on the contrary, we want to find out how experienced teachers interpret and use the materials.

We will use at least three indicators to measure effective teaching: (1) the ability to carry out the instructional activities in highly rated materials, (2) the ability to adapt materials for new contexts and needs, and (3) improvements in student learning. Indicators of teacher knowledge will include teachers’ knowl-
edge of (1) learning goals and alignment of materials to them, (2) instructional analysis criteria and how to apply them, and (3) alignment of assessment to learning goals.

Data collection will consist of use of teacher logs/journals as well as classroom observations and teacher interviews. Researchers will observe or videotape a teacher’s class approximately twenty times. In some classrooms, we will use Polycam technology to observe and/or videotape lessons or to carry out interviews with teachers or students. Polycam provides remote one- or two-way video and audio over the web, with the capability to control remotely the direction of the camera. At least one additional contact will be made each month, with informal interviews and review of teacher logs as needed to maintain a reasonable level of description of classroom teaching between observations. More intensive data collection may be done in selected cases as resources allow.

Data about the enacted curriculum will be entered into databases parallel to those for the literal program. Notes and predetermined descriptors will be entered for each task students are to carry out, such as listening to teacher presentations, reading text materials, making predictions, collecting data, and discussing results. Audio or video recordings will enable observations. Protocols for classroom observations and videotaping will be adapted from other research projects (e.g., Smith & Anderson, 1983; Salish, 1997) and/or will be created for this research.

The enacted curriculum will be compared to the literal program. Patterns of modification will be identified, including the kinds of tasks that are omitted, added, or modified. Of particular interest will be how the modifications relate to the criteria for good instruction that are identified in the Project 2061 curriculum-materials analysis procedure. Criteria in categories II and V provide a basis for distinguishing among materials and may be helpful in distinguishing among teaching practices. (See Appendix I for sample criteria, indicators, and rating scheme.) In subsequent analyses, the patterns of use and modification of the recommendations of the curriculum materials will be related to patterns of student achievement.

Issues of validity and reliability of classrooms observations are key for this research. To ensure reliability we will train graduate students and mathematics specialists to do the classroom observations. Using videotapes of teaching, we will follow protocols reported in the literature for video analysis and ensure that observers use the instruments in standard ways (e.g., the training should include at least three people watching the tape together and then sharing their observations) (Gallagher & Parker, 1997; Schoenfeld, 1992). The data will be collected in the different sites where the research is being done. Protocols for analysis will be developed by the researchers under the leadership of Project 2061 to assure that the same standards are used in the different sites for collection and analysis. Both data and analysis will be collected by Project 2061 to be used by the analysis team.

**Assessments of student learning.** Student learning in science and mathematics is not just accumulation and repetition of knowledge. Research from cognitive psychology and from investigations of expertise and of students’ misconceptions (and alternative conceptions) makes a persuasive case for cognitive restructuring of knowledge that accompanies the development of expertise (Chi et al., 1981; NRC, 2000). In our view, student learning should be a coherent web of understanding. The nature of that web is described in the conceptual-strand maps in *Atlas of Science Literacy* (see Appendix C). One goal of understanding change in student learning will be to characterize cognitive restructuring of students’ knowledge. In particular, differences in slopes of learning curves of individual students are an important sign of instructional effectiveness. Such research is inherently intra-individual and requires assessments sensitive to such changes.

The measures of student learning will address several interrelated factors: (1) the alignment of assess-
ments with learning goals—both content alignment and the likelihood that tasks effectively assess the targeted goals, (2) the formality of measures—standardized and informal, (3) the balance of assessment tasks—multiple choice and constructed response, and (4) the unit of reporting on achievement—classroom and individual student.

Alignment of tasks with learning goals. Project 2061 has developed a procedure for analyzing the alignment of individual science and mathematics assessment tasks with specific learning goals, such as statements from national or state standards (NSF no. ESI 98-19018). The procedure addresses both the alignment of content (Is knowledge of the goal both necessary and sufficient for responding to the task?) and the likely effectiveness of the task (Is the task clear and fair in its statement and expectations, engaging and interesting to students, and not susceptible to guessing?). (See Appendix J for a description of the procedure.)

Formal and informal assessments. The states involved in the study have in place standardized mathematics assessments, and some districts use standardized tests as a regular component of their accountability system. These state assessments or standardized tests are central to school and teacher decision making about the curricula they use and the learning goals on which they focus. For this reason, the project will use these tests as one measure of student achievement. The tests offer an opportunity to report achievement on standardized measures, making it possible to compare student achievement in the project with previous years in a particular school or with similar schools in the state. On the other hand, most standardized tests are comprised primarily of multiple-choice items and may not probe the targeted learning goals. The study will also incorporate (1) specially designed assessment tasks aligned with the learning goals and (2) a range of informal assessments, including homework and quizzes, teacher judgments, student interviews, hands-on tasks, and student projects to help interpret nuances in individual or classroom achievement.

Balance of assessment tasks. Balance is essential in the interest of fairness and to provide an opportunity to assess the best achievement and learning for all students. In addition to various types of tasks (e.g., multiple choice, short response, extended response), the study will use tasks that incorporate different response modes (e.g., written, verbal, performance, ESL) and a range of contexts and applications, levels of novelty, and technology or tool support.

Development of a pool of tasks. Specifications for the assessment instruments and the number and types of tasks needed to construct them will be developed. The pool of tasks needs to be sufficiently large to allow for some attrition as the tasks are piloted. All candidate tasks will then be evaluated using the draft procedures developed by Project 2061. Tasks that pass the evaluation or can be revised to do so will become part of the task pool. Tasks for the pool will be sought from the assessments for each of the materials to be investigated, from released items from the National Assessment of Education Progress (NAEP) and the Third International Mathematics and Science Study (TIMSS), from state assessment programs, and from standards-based assessment projects. Additional tasks will be developed or adapted by project staff as needed to complete the pool. Estimated size of the task pool: 4 specific learning goals x (5 constructed response + 8 selected option tasks + 2 performance tasks per goal) = 60 tasks.

Task piloting and revision. Tasks in the pool will be piloted with representative samples of 10-12 students who will also be interviewed about their thinking as they responded to the tasks. These results will be used to evaluate and, where necessary and feasible, revise the tasks. The revised tasks and response features will be used in task tryouts with larger representative samples of 200-250 students. Analysis of student response data will examine the reliability of response feature coding and patterns and the relationships among tasks for the set of learning goals that will be used to assess construct validity. The draft instruments will be administered to samples of approximately 500 students as part of a pilot study of the achievement of students in classrooms where highly rated and lower-rated materials are being used. Samples of 50 students, stratified on the basis of their overall performance, will be interviewed to deter-
mine concurrent validity. For example, high/medium/low classifications of assessment task responses will be tabled against similar classifications from interviews. Agreement between the interviews and the use of the assessment instruments will be evaluated. Minor revisions of tasks may be made, but major revision or replacement of tasks would require additional empirical testing.

**Monitoring ongoing support.** The primary purpose of gathering data on ongoing support is to determine whether the communities of practice established during the first three years of the longitudinal study continue and whether new teachers are inducted into them. We will look to see whether time continues to be made available for groups to meet and whether the kinds of discussions about teaching and learning specific ideas continues, whether it includes new teachers, whether it expands to new ideas, whether it includes new data collection, and so forth. We will also observe the extent to which the rationale for maintaining these discussions is communicated to teachers and administrators, parents and policymakers.

**Timetable.** A year-by-year description of the longitudinal study and the experimental study is provided below (see Appendix E for display of the components and sequencing of the studies):

**Year 1** will serve as a pilot test of our methodology and instrumentation and provide baseline data for studies in the subsequent years. Several different sets of teachers will continue to use the textbooks they are already using and which have received high, moderate, or low ratings in supporting effective instruction (AAAS, 2000). We will examine teachers’ patterns of implementation and student achievement on the targeted learning goals.

**Years 2-3** will focus on what additional professional development is needed to improve the quality of teaching, the implementation of the materials, and student achievement of the targeted learning goals. The professional development will emphasize understanding of research-based teacher practices, how highly rated curriculum materials support teachers, and how the use of these materials relate to student learning.

**Year 4** will be a continuation of the longitudinal study of the Year 1-3 teachers, now focusing on persistence and sustainability. In addition, we will use revised professional development and technological support in the experimental study of completely different sets of veteran teachers in the same sites, again using the textbooks that were used in Years 1-3.

**Year 5** focuses on the Year 4 cohorts of teachers who continue to use their materials. This will allow us to judge the adequacy of the technology support with optimal compared with minimal professional development. We will also introduce the Year 4 cohorts to a new set of topics for which they have not received specific professional development. This will provide an opportunity to study how transferable their new teaching practices are and to measure the impact on student learning. There will be no additional professional development for Year 5, but the technology-based support will continue. We will continue to examine the sustainability of implementation with the Years 1-3 cohorts of teachers.

**IERI Benchmarks Addressed**

This proposal addresses the four IERI benchmarks in the context of Focus Area II: *Transition to Increasingly Complex Science and Mathematics Learning*. We study promising educational practices that can increase students’ learning of mathematical ideas as they move from middle to high school.

**Benchmark I: Research methodology.** We have combined the power of statistical models with the richness of qualitative data to make sure that we obtain valid, reliable, and useful data. Statistical models used in this proposal include structural equation modeling and hierarchical linear modeling, as described in the Research Methodology section. Qualitative methods include semi-structured interviews, extended
observations of teaching in classrooms, analysis of videotaped teaching, and student work samples. Our longitudinal interventions will provide data to construct growth curve models of student learning that can be associated with the classroom use of highly rated curriculum materials and professional development tailored to effective research-based teaching practices. Although the selection of the samples is based on what is available (naturalistic conditions), the samples represent schools and districts that have teachers and students with different backgrounds and SES levels. We will develop and test instruments for assessing teacher use of materials and student understanding of mathematics. Reliability and validity of these measures will be established the first year of the research and will be continually checked.

**Benchmark II: Scalability.** To scale up promising practices we need to be able to transfer research knowledge into school settings. Establishing the scalability of an intervention requires establishing that (a) these practices work on a small scale, (b) they can be sustained at least at the same scale for some period of time in the original cohorts, and (c) they can be extended to other cohorts and transferred to other content. The first is addressed through small-scale studies in Years 1-3; the second, in Years 4 and 5 (same cohort); and the third, in Years 4 and 5 (new cohort and new content). Note that there are at least three interrelated aspects of scaling up in our proposal. One entails examining whether practices can be extended to more teachers. A second involves examining whether the original cohorts of teachers are able to teach content different than that on which the professional development interventions were based. And a third aspect of scalability involves the use of technology to increase the efficiency of providing the professional development interventions. An underlying issue is the relevance of these interventions to diverse learners. The seven school districts involved in the study serve diverse populations of students and teachers to the extent that all of our interventions will be tested with representative samples of low SES and minority students.

**Benchmark III: Technology.** In this research we use technology as a means to collect data on the research itself and as a fundamental support to sustain and scale up effective teaching practices that improve student learning. As part of the data collection, researchers will observe or videotape a teacher’s class using Polycom technology to observe and/or videotape lessons or to carry out interviews with teachers or students. Polycom provides remote one- or two-way video and audio over the web, with the capability to remotely control the direction of the camera. We will use the Internet to make available teacher support materials, including, for example, student work samples or video clips of teachers carrying out instructional activities with helpful comments about implementing them. Additionally, a special system of “criss-crossing” of parts of the video cases, based on cognitive flexibility theory, will be used to deepen understanding of the cases and to accelerate the acquisition of interconnected representations of experience. In doing so, we will document the potential that technology has for supporting, sustaining, and scaling up effective teaching practices.

**Benchmark IV: Interdisciplinary research team.** We have assembled an interdisciplinary team (see Appendices K, L, and M) to design and execute the proposed project. The team includes cognitive scientists, mathematicians (including statisticians), information technologists, education researchers and teacher educators, educational policymakers, and K-12 teachers and curriculum specialists. Collectively, they bring expertise in qualitative and quantitative research methodologies (V. Willson, Ahlgren). They are national and international authorities in developing and clarifying standards (Ahlgren, Roseman), evaluating curriculum materials in light of standards (Roseman, Kulm, Morris), analyzing teaching (Hiebert, M. Smith, Roth), designing assessment aligned with standards (Kulm, L. Wilson, Manon, E. Smith), and in considering policy issues of professional development and teacher learning (Sykes, Rosen). They bring expertise in telecommunications and information technology research and their application to cognition and learning (Spiro, Striker). In addition, team members Hollowell and Kulm have excellent long-term relationships with local schools and districts that will be fundamental for the
success of the proposed research. (See letters of support in Appendix A.) In addition to the NBPTS-
certified teachers and evaluators on the Advisory group, we view the fifty participating teachers as full
partners in the research. The leading role of AAAS Project 2061 will ensure that continual input and
interaction will take place during this research ensuring common data and analysis. (See Appendix N for
a list of responsibilities.)