Research Issues in the Improvement of Mathematics Teaching and Learning through Professional Development

George DeBoer, Kathleen Morris, Jo Ellen Roseman, and Linda Wilson
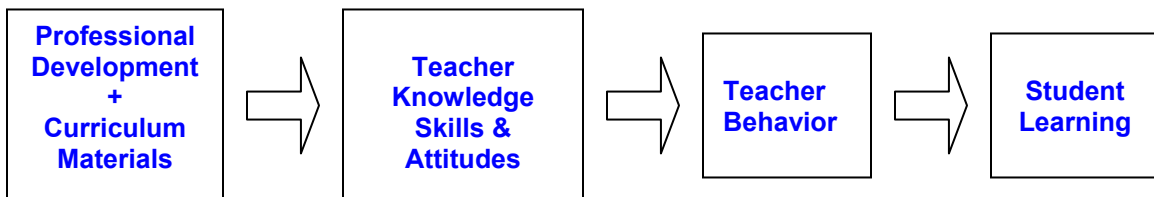**American Association for the Advancement of Science**

Mary Margaret Capraro, Robert Capraro, Gerald Kulm, and Victor Willson
Texas A&M University

Jon Manon
University of Delaware

The purpose of this paper is to describe a study we are conducting on the improvement of mathematics teaching and learning at the middle school level through professional development and to discuss some of the research issues that we have encountered in conducting the study. The paper will lay out the various rationales for our initial design and for the adjustments that we made along the way. We are nearing the end of year two of a five-year study, so this is very much a work in progress. The study is not large in terms of the number of teachers involved (approximately 50 teachers and 1,000 students per year in the early stages of the study), but it is a complex study involving many interconnected elements. In Part I we lay out the design of the study, and in Part II we discuss some of the issues that we are facing as we progress through our work.

Part I: The Design of the Study

The purpose of this IERI[1] project is to learn how to provide, on a large scale and making use of Web-based and video technology, the professional development and continuing support that teachers need to improve student achievement in mathematics. The study is led by a team of researchers at Project 2061 of the American Association for the Advancement of Science, Texas A&M University, and the University of Delaware. Participating teachers and their students are from school districts in Texas and Delaware. The study involves investigating the relationships among professional development, curriculum materials, teacher knowledge, teacher behavior, and student achievement. The logic model is presented below:

| Professional Development + Curriculum Materials | → | Teacher Knowledge Skills & Attitudes | → | Teacher Behavior | → | Student Learning |

The logic model is presented as an oversimplified representation of the linear relationships that we are testing. As such, it leaves out the many feedback loops that also exist, such as the effect of student learning on teacher behavior and attitudes and the impact of teacher attitudes and knowledge on professional development.

The study is being conducted in two phases. In the first phase, we are creating a classroom observation utility and instruments for assessing student learning, exploring relationships among the variables in the study, and experimenting with the content and the modes of delivery of professional development. This is very much an exploratory phase during which many of the issues raised later in this paper will be resolved. In the second phase, we will involve a larger number of teachers in comparing the effectiveness of several models of professional development using random assignment of treatments to subjects.

We began with a number of key assumptions and decisions that guided our work:

(1) Learning Goals. In planning this study, we designed it around three specific learning goals for students. This is consistent with current thinking in mathematics education, where a coherent picture of what all students should know and be able to do by the end of their K-12 education is defined through the learning goals set forth in content standards (National Council of Teachers of Mathematics [NCTM], 2000). Designing the study around specific learning goals also enables us to maximize the precision of our measurement of student learning and sharpen the focus of our classroom observations of teachers. The three learning goals are in the areas of number and operations, algebra, and data analysis. These learning goals were chosen because of their central place in both national and state content standards. With minor differences in wording, each of the three is represented in *Benchmarks for Science Literacy* (American Association for the Advancement of Science [AAAS], 1993), *Principles and Standards for School Mathematics* (NCTM, 2000), and in the state standards of both Texas and Delaware. The three learning goals are:
- Use, interpret, and compare numbers in several equivalent forms such as integers, fractions, decimals, and percents.
- Symbolic equations can be used to summarize how the quantity of something changes over time or in response to other changes.
- Comparison of data from two groups should involve comparing both their middles and the spreads around them.

It is important to note that two of the learning goals identify content knowledge and one identifies skills to be mastered. We take the position that regardless of the form of the original statement, skill implies understanding and understanding implies the ability to use and apply the knowledge. In some cases it is more appropriate to write the learning goal in the form of a knowledge statement and in other cases it is more appropriate to write it as a skill statement. In neither case should this be viewed narrowly as skill without understanding or understanding without the ability to apply that knowledge.

(2) Curriculum Materials. Because curriculum materials play such a large part in classrooms by defining the content and how it is taught, learning goals and pedagogy

cannot be considered separately from the materials that are used. In most classrooms the textbook is the primary vehicle through which learning goals are communicated to students and through which the intent of the developer is communicated to the teacher. These materials have the potential to serve as sources of reflection for teacher learning as well as for student engagement with important mathematical ideas (Acquarelli & Mumme, 1996; Ball & Cohen, 1996; Smith, 2001). The interplay between the teacher and the materials is, therefore, a very important element in what ultimately becomes the enacted curriculum (Zumwalt, 1989). (Of course, other factors, such as the interactions between student and teacher and student and materials are also essential parts of what becomes the enacted curriculum.) In addition, the approach to professional development that we took is one where teachers learn about the extent to which their textbooks do or do not support certain pedagogical practices. In this way, the materials ground the professional development in the real-world practice of teachers. Thus, there is an intimate relationship between curriculum materials and pedagogy that must be considered in all aspects of the study.

The textbooks that were included in this study vary in the amount of pedagogical support they give teachers. We intentionally included a variety of textbooks because we wanted variation in the way teachers met or did not meet the instructional criteria on which we were evaluating their teaching. Of the teachers we chose for the study, some are using materials that had scored well on the Project 2061 textbook evaluations (AAAS, 2000), some are using materials that had received a mid-range score, and some are using materials that scored at the low end of the range. The materials being used are *Connected Mathematics* (Lappan, et al., 1998; Lappan, et al., 2000), *Mathematics in Context* (Romberg, et al., 1998), *Middle Grades Math Thematics* (Bilstein, et al., 1999), and *Mathematics Applications and Connections* (Collins, et al., 1998). Involving students from two different parts of the country, Texas and Delaware, also serves to increase variability on our measurements and the generalizability of our results.

(3) Defining and Measuring Quality Teaching. As with learning goals for students, we also believed that it was important to define quality teaching in a very precise and measurable way. The criteria we chose were the ones that Project 2061 had used in its evaluations of mathematics and science textbooks (AAAS, 2000; Kesidou & Roseman, 2002). Although it was necessary to modify some of the criteria so they could be used to describe teacher behavior, we found that most of the criteria for evaluating textbooks could be used without adaptation to evaluate teaching as well. In all, there are seven categories encompassing 24 criteria and approximately 125 quality indicators.[2] The seven categories are:

1. Providing students with a clear sense of purpose for what they are learning.
2. Taking students' existing ideas into account.
3. Engaging students with phenomena and real-world examples.
4. Helping students understand new ideas and how to make use of them.
5. Promoting students' thinking about what they are learning.

---

[2] A complete list of criteria and quality indicators can be found at the Project 2061 Web site, www.project2061.org.

6. Assessing student progress to provide feedback to students and teachers.
7. Enhancing the learning environment for all students.

It should also be noted that meeting the instructional criteria means meeting them in the context of the learning goals. That means, for example, that when we look at the seven categories of instructional criteria, each one needs to be viewed as if the phrase "related to the learning goals" is included. To illustrate using the first three categories:

1. Providing students with a clear sense of purpose for what they are learning [*related to the learning goals*].
2. Taking students' existing ideas [*related to the learning goals*] into account.
3. Engaging students with phenomena and real-world examples [*related to the learning goals*].

To determine if there is any independent effect of content alignment, separate from instructional quality, we also are measuring the amount of time spent on each learning goal. Perhaps simply introducing students to the language of the learning goal has a positive effect on test performance. Or, conversely, it may be that content alignment alone is of little value if it is not accompanied by quality teaching. If so, this will significantly affect the way we speak about content alignment in teaching.

For each of the three learning goals on which the study is focused, we observed three to five lessons for each teacher. The lessons chosen were the ones most closely related to the target learning goals and the ones that give us the best chance to apply the instructional criteria. The lessons were videotaped and then analyzed at a later date. The videotaped lessons were also used as a tool in professional development during the second summer and will be a significant part of the professional development models that will be tested in the second phase of the study. Along with the curriculum materials that the teachers are using, the videotaped lessons will help ground the professional development in the real-world experiences of teachers.

A Web-based computer utility was developed to provide structure and serve as a portal for our analysis of the curriculum materials (textbook analysis) and instructional quality (instructional analysis). Each videotaped lesson is stored on a CD-ROM and used in conjunction with the utility. Videotaped lessons are time-coded and evaluated to determine the extent to which the content of the enacted lesson is aligned with the learning goal and with specific instructional criteria. All of the study data will be captured and stored on the utility.

(4) Professional Development. This study is ultimately about the extent to which teachers, who are supported in varying degrees by the curriculum materials they are using and by professional development, change their instructional patterns so that they become more goal oriented and develop competence in pedagogical practices that lead to student understanding of the target learning goals.

The professional development workshops are being designed following the same research-based instructional practices that we want teachers to use in their classrooms.

Just as we think that teachers' use of these practices (in the context of specific learning goals) will increase their students' learning, so we think that our use of these practices (in the context of our learning goals for teachers) will increase teacher learning. Whether our learning goals for teachers have to do with understanding the mathematical ideas in the student learning goals, selecting and using representations to clarify the ideas for students, selecting and using questions to guide student thinking about the representations or probing their understanding of the ideas, the professional development workshops will:

- Provide a sense of purpose for teachers based on the demonstrated needs of their students.
- Take account of teachers' initial ideas and skills.
- Engage teachers in relevant classroom phenomena such as the results of student learning, videotapes of their own teaching, and the relationship between their teaching and their students' learning.
- Help teachers develop and use ideas and skills through ongoing practice and support.
- Promote teachers' thinking about their experiences with the phenomena of the classroom through probing questions.
- Assess progress and offer feedback to both teachers and professional development providers.
- Enhance the professional development learning environment by encouraging a spirit of healthy questioning and by enabling all participants to experience success.

The professional development takes a case study approach. (For a discussion of the use of case studies in mathematics education, see Lampert & Ball, 1998; Shulman, 1992; Stigler & Hiebert, 1999.) Teachers are (1) introduced to the instructional criteria through examples and counter examples from their materials and from videotapes of their teaching, (2) given practice applying the criteria to new examples while receiving feedback from the leader and peers, and (3) expected to apply the criteria to their own teaching. In each workshop teachers carefully examine a subset of the instructional criteria, learn how to recognize from videotapes and their textbooks when those instructional criteria are evident in lessons aimed at particular learning goals, and consider how best to use their textbooks effectively to apply this knowledge and these skills in their own teaching.

In a two-day session during the first summer, teachers were given an orientation to the project and were introduced to the learning goals on which the study is based. They were shown the wording of the learning goals from Project 2061's *Benchmarks* and the similarities between that wording and the wording of their own state standards and the NCTM *Standards*. They also spent time interpreting the meaning of each learning goal, examining the research on student learning (including prerequisites and relevant student misconceptions) for each learning goal, and discussing what their analyses and study of the research implied for instruction and assessment.

Focusing the first professional development workshop on learning goals was done for two reasons. First, we wanted teachers to understand from the start what it means to focus

their instruction on a learning goal. Second, we expected that focusing first on the meaning of the learning goals, on prerequisites, and on relevant student misconceptions associated with the learning goals would help increase teachers' understanding of the content. To that end, we also engaged teachers in considering how well various lessons were aligned with the learning goals. The task of analyzing the alignment of lessons gave them a chance to apply (and further test and improve) their understanding of the learning goals. Focusing on the learning goals was also necessary preparation for the second phase of professional development that took place the following summer when teachers considered a number of instructional criteria with respect to the learning goals.

Prior to the second summer, a decision was made to focus initially on three instructional criteria: (1) using representations effectively, (2) probing student understanding, and (3) guiding student interpretation and reasoning. Our choices grew out of what we were observing on the videotapes, what improvements we thought would have the largest impact on learning, and what we thought we could accomplish in five days of professional development during the summer. (The total number of professional days throughout the entire year is ten, but only five are being used in the summer.) Although we had not completed a systematic analysis of all the videotapes, our initial examination across teachers and textbooks revealed some notable patterns of practice that needed improvements:

- Teachers were often asking questions to find out if students had gotten the right answer to problems rather than probing students' initial ideas and monitoring the development of their conceptual understanding.
- Teachers were not systematically using representations to help students make sense of abstract ideas. Instead, when students got stuck on problems, teachers would often ask even more abstract questions rather than focusing students on representations they could use to figure the problem out.
- Teachers were asking guiding questions that focused on helping students complete procedures correctly and on getting the right answers rather than on building students' conceptual understanding.
- Teachers were not attending to the sometimes diverse learning needs of their students. (Videotapes of small group work revealed that some groups would complete a task quickly and have nothing to do while other students were having difficulty with the initial task.)

In short, teachers were not focusing their instruction on building students' conceptual understanding of the learning goals. Focusing professional development on representations and questioning strategies seemed a potentially fruitful way of moving them in that direction effectively.

Of the three instructional criteria that we focused on in the second summer, we chose to start with using representations rather than on strategies for "probing student understanding" or "guiding student interpretation and reasoning." We did this for several reasons. First, representations would keep the content of the professional development grounded in the learning goals, something that was proving to be difficult for teachers to do in their classrooms. In examining representations in textbooks and teaching, the

professional development activities would consider (1) what idea was being represented, (2) whether it was being represented accurately, (3) whether it was likely to be comprehensible to students, and (4) whether instruction engaged students in considering what aspects of the idea were being represented and what aspects were not. This would situate the professional development in the ideas that students were supposed to be learning and in the context of how teachers were interacting with students in the classroom. Second, we thought it was necessary for teachers to have a solid grounding in the mathematical ideas themselves and the ways of representing those ideas before they could think productively about how to improve their strategies for questioning students. Third, representations are the primary instructional tool that textbooks use to make abstract ideas clear to students. All the textbooks included some helpful representations. Teachers would likely be familiar with these representations and would welcome the opportunity to examine others that might help particular students. Moreover, examining a range of representations from the other textbooks would help teachers identify critical attributes of representations that are helpful for clarifying abstract ideas. Finally, we had identified video excerpts of students having problems using representations that we thought would be compelling to teachers.

(5) Teacher Knowledge and Attitudes. We expect that the knowledge and attitudes that teachers come with, both of mathematics and pedagogy, will have an impact on what they learn from professional development. The knowledge and attitudes they acquire during professional development should, in turn, affect their behavior in the classroom. For example, it is expected that their knowledge of mathematics may impact how successfully teachers are able to understand and teach a curriculum that focuses on promoting a deep understanding of mathematics in their students (Cohen, McLaughlin, & Talbert, 1993; Fennema & Frank, 1992; Hiebert & Carpenter, 1992). Likewise, a positive attitude toward the proposed pedagogies should affect how willing teachers are to change what they are currently doing. These attitudes that teachers initially have may, in turn, change as an outcome of professional development.

To measure prior attitudes and knowledge, we administered a baseline survey at the beginning of the second summer. Teachers were asked to indicate: (1) high school and college mathematics courses they had taken; (2) number of years they had taught mathematics and the number of years they had taught mathematics at this level; (3) number of hours of professional development they had completed related to their textbook, to mathematics, and to topics of a more general nature; (4) courses they had taken in mathematics education and in non-mathematics education; (5) level of preparation they felt they had to engage in various classroom teaching practices; and (6) support they had received in those teaching practices from their textbooks, from previous professional development, and from formal coursework.

To measure teacher knowledge and attitudes as outcomes of professional development, we are considering asking teachers to complete a number of open-ended tasks that will allow us to estimate how much they have learned about the ideas being presented and, perhaps, give us some insights into their attitudes as well. In the example below, the first question deals with teacher understanding of the instructional criteria that were the focus

of the professional development. The second question deals with teachers' understanding of content alignment and their ability to recognize and evaluate the extent of content alignment in their textbooks:

> Instructions to Teachers: To help us further refine our instructional criteria and to help textbook developers incorporate them more effectively into curriculum materials, we ask your help in the following ways. Please think carefully about the instructional criteria that were highlighted this week and then offer your judgments about what is most important about the criteria and how they might be improved. Think also about the learning goal we focused on and how well your textbook addresses that learning goal.
>
> 1. How well do you think the textbook you are using supports you in implementing these instructional criteria? Give some examples of what it is that the textbook does to help you implement these instructional criteria. How effective are the activities that the textbook uses, that are related to these instructional criteria, for your own students? How could they be improved?
>
> 2. How well do you think the textbook you are using addresses the learning goal that we have been focusing on during this workshop? Is the content specified in this learning goal presented accurately in your textbook? Is it presented coherently? Is the content aligned with the learning goal? Give examples of where these things are done well and where they could be improved.

These questions are of the analysis/evaluation type. Teachers would be asked to complete these tasks as a professional activity to provide the research team with feedback on the instructional criteria and how well they are incorporated into their textbooks.

We are also experimenting with other ways of obtaining information about teacher knowledge and attitudes. For example, at the end of the second summer's workshop, we asked teachers to write a lesson plan incorporating some of the ideas they had learned. We also plan to have teachers use the instructional criteria and the Web-based utility to analyze videos of their own teaching. These tasks should provide information on how well teachers understand the ideas presented in the workshops. We do not plan to measure changes in their understanding of mathematics as this is not a primary focus of this study.

(6) Student Assessment. The aim of the assessment instruments used in this study is to measure student understanding related to the learning goals. To ensure precision in the measurement of student understanding, the learning goals were first analyzed by a group of experts in mathematics education to determine the ideas that comprise the learning goals. The analysis entails studying in detail the language and intent of each learning goal within the context of the relevant research on student learning (including research on prerequisite knowledge and common misconceptions) and defining the precise set of mathematical ideas included in that learning goal. The process produced a list of key ideas for each learning goal. For example, for the Data learning goal: "By the end of 8th grade, students should know that comparison of data from two groups should involve

comparing both their middles and the spread around them," a set of seven idea statements were produced. Specifically, students who achieve this learning goal should be able to:

1. Organize data into meaningful units.
2. Read and interpret various representations of a set, or sets, of data.
3. Select and construct an appropriate representation for a set of data to illustrate characteristics of central tendency or dispersion, such as box plots, line plots, and stem and leaf plots.
4. Understand concepts of mean, median, and mode. Recognize the effect of additional data points on measures of central tendency.
5. Determine the mean, median, mode, and range for a set, or sets of data, presented in various forms.
6. Select and use an appropriate measure of central tendency to describe a set of data, including mean, median, and mode.
7. Describe the spread of a set of data, using measures such as range and quartile and descriptors such as outliers, clusters ("clumps"), and gaps.

The grain size of these skills and ideas was determined by considering what could be assessed in the given format by a small number (approximately one to five) of test items. The statements were then organized into broader categories. In the case of Data, statements 1through 3 above were placed in a category called "Representation," and statements 4 through 7 were categorized as "Summary Measures." An assessment map for this learning goal is shown in Figure 1.

The assessment map became the basis for the development of all test items and for the distribution of items selected for the assessment instrument. Similar maps were developed for the Number and Algebra learning goals. For each learning goal, approximately eight items were in a multiple-choice format, and five to eight items were in a short constructed response in format. Short constructed response items ask students to generate a single answer, an answer with a brief explanation, or a display of their work. In addition, a single extended constructed response item, consisting of four to five scaffolded parts, was developed. This item, designed as a "super item" (Collis, Romberg, & Jurdak, 1986), was used to "anchor" the instrument and was intended to be a direct measure of the learning goal rather than of its component parts. The multiple choice and short response items were used to assess the component parts of the learning goal, including any prerequisite knowledge or misconceptions.

Some of the test items are situated in a real-world context while others are not. The number of contextualized items varies across the learning goals. The Data test, for example, consists almost entirely of contextualized items, given that the learning goal deals with comparing data, which arises from contexts of various kinds. The Number learning goal, on the other hand, deals with using, comparing, and interpreting numbers in different equivalent forms, and a smaller percentage of that test is made of contextualized items. When contexts were used, they were chosen to enhance students' interest, engagement, and motivation or to measure students' ability to apply concepts or procedures in a variety of real-world situations.

All of the test items were piloted with middle school students who were not otherwise involved in this study. These initial data were used to revise or delete items that were in development. The constructed response items, both short and extended, were also piloted using cognitive labs in which students were interviewed one-on-one immediately after completing the item. The purpose of the cognitive labs was to determine the clarity, accessibility, and content validity of the items, and items were revised or deleted accordingly. Prior to field testing the items, the open response items were examined by analysts who had been trained in Project 2061's procedure for analyzing the alignment of assessment items to learning goals (AAAS, 2004). All items were then field tested in non-study classrooms that were determined to be comparable to the classrooms of participating teachers. Final revisions were made following the field testing. Parallel forms of the items were created for post testing by modifying the context of the "super item."

Reliability and Validity. For the constructed response items on the Algebra and Number tests, inter-rater agreement for the pre- and post-tests ranged between 94 percent and 99 percent. Cronbach Alpha reliability scores for the first year of post-tests were .72 for the Number test and .82 for the Algebra test. Factor analyses are now being conducted to obtain information on the factor structure of the tests. Reliability measures were based on the responses of over 1,000 students at each administration of each test. To ensure item validity, test items were constructed to match the learning goals. Judgments concerning the alignment of items to the ideas in the target learning goals were made by experts in the field who used an alignment procedure developed by AAAS Project 2061 (AAAS, 2004) and by the use of cognitive lab interviews to determine if students' responses on test items matched the understanding they demonstrated in the interviews. During the process of item development, items that were judged not to be aligned, based on results from the expert analysis or from the cognitive interviews, were discarded or revised.

Part II: Research Issues

In carrying out a project such as this, a number of theoretical and practical issues arise, some of them anticipated and some of them unforeseen. Many issues result from trying to conduct research in the real world of schooling. Inevitably the influence of local, state, and national level contexts becomes important to the success of the project. Working within already established frameworks of policy, beliefs, and professional practice offers both challenges and opportunities. States have content standards and testing programs. They must meet the requirements of the federal education legislation, No Child Left Behind (2002). States and local school districts vary in their requirements for teacher professional development and the organizational support that is provided, including how much and when time is available for professional development. Districts vary in the average length of time teachers remain in the district and how frequently teachers' assignments change.

State-level mathematics assessments have had a particularly powerful influence on the educational system. State assessments are used to determine individual students'

progress and provide accountability measures for teachers and administrators. It is difficult, for example, to overstate the influence of the Texas Assessment of Knowledge and Skills (TAKS) on schools and teachers. Literally every decision that is made with regard to curriculum selection, scheduling, and professional development is driven by the requirement to improve students' scores on the TAKS. In the case of schools that have been identified as low performing, the pressure to improve scores can drive out any program or activity that is perceived not to contribute directly to the goal of raising test scores.

Although this study was carefully designed with these contextual issues in mind, the pervasiveness of their influence cannot be overstated. As we progress through the study, we are continually faced with making decisions and adjustments, many of which involve having to accommodate the realities of schooling. Others arise from unresolved theoretical issues that will be worked out during the course of the study. In this section of our paper, we discuss four key areas in which issues and questions regarding the research have arisen and how we are proposing to deal with them. The four areas are: (1) securing and retaining teacher participants, (2) defining and measuring quality teaching, (3) professional development, and (4) student assessment.

(1) Securing and Retaining Teacher Participants. Teacher turnover has been a major concern for us in the study. (The overall return rate for teachers in our study is 71 percent.) This has affected sample size and, therefore, some of the questions we can answer. Using replacement teachers has implications for the design and delivery of professional development and has presented challenges for analyzing and interpreting outcome measures. As originally conceived, the first three years of our work were to be devoted to a longitudinal study in which a cohort of teachers would receive professional development incrementally over those three years. We would observe changes in their teaching and in their students' achievement along the way. In the second phase of the study, we would use baseline data gathered in the previous year to measure the impact of professional development the next year. So at a minimum, in the second phase of our study, we will need teachers who will stay in the study for two years.

There are a number of reasons why teachers leave the study. They may leave teaching temporarily or permanently for personal reasons or they may be transferred by the administration to another school or to another assignment within the same school. Because their participation in the study is voluntary, they may leave because they do not feel that the time they spend will positively affect their students' learning. Administrators who have doubts about the value of the study may also encourage teachers to drop out. Although we obtained commitments from teachers and administrators at the beginning of the project, we have experienced a significant amount of turnover already. To reduce turnover, here are some of the things we now do or are considering doing:

- **Develop collegial relationships with participating teachers.** In this study, teachers are treated as part of a team of professionals engaged in important work. At the beginning, we held informational meetings, often over a meal, to inform teachers of the project goals, their roles in the project, and their expected level of

commitment. We emphasized the research goals and the importance of the teachers' part in advancing knowledge. A monthly newsletter provides information about the other partners and highlights the work of one of the teachers, especially focusing on their professional achievements such as presentations at conferences or publications of their ideas.

- **Ensure the relevance of the study to the teachers' work.** In a pilot study conducted in Texas prior to the project, for example, performance on the TAKS was a key to gaining interest and participation by the teachers (Kulm, Capraro, Capraro, & Hastings, 2002). Teachers must believe that the effort they are expending in the study will have payoffs in terms of their students' achievement on their state assessments and will not detract from that success. To this end, teachers helped us to identify the key ideas in the learning goals at each grade level that were most important for their state test. Teachers came to believe that the tests used in the study would not only assess students' understanding of the learning goals more thoroughly than the state tests do but would also be more helpful in diagnosing their students' learning difficulties because of the greater number of items that are focused on the target learning goals. The teachers were convinced that the study tests could provide feedback about their students' achievement that would be specific enough for instructional redesign.

- **Encourage senior teachers and lead teachers to participate.** Peer influence can have powerful effects on participation, especially on beginning teachers. If the senior teachers in a school choose not to participate, it may be difficult for less-experienced teachers to continue, especially in cases where a new textbook and teaching strategy are being implemented. Teachers who have standing within their schools should be recruited to maximize peer influence.

- **Recruit teams of teachers from each school.** The participation of two or more teachers in a school allows for cooperative planning of lessons aligned to the learning goals and provides teachers with a colleague with whom to analyze videotapes of their lessons. Groups of teachers should be recruited to enhance opportunities for collaboration and support within a community of practice.

- **Enlist the support of key administrators.** District administrators, school principals, mathematics supervisors, and master teachers must all recognize the importance of the professional development the teachers are receiving and the changes that are expected in their teaching. The underlying influence of state assessment is felt at all of these levels, where any new approach or activity is weighed in relation to its potential impact on students' test scores. Key administrators must be convinced of the value of the study for teachers.

- **Build on past collaborative efforts.** Each university has established its reputation through previous collaborations and through personal contacts with school leaders. These good relationships are often the primary means of encouraging and maintaining participation of teachers. Often teacher participants begin graduate work during the project or join the project as participants after completing certification at the university. Building on these previous efforts and maintaining the personal relationships between schools and universities is an important way to secure and extend participation.

(2) Defining and Measuring Quality Teaching. Although our intent is to analyze the teachers' videotaped lessons with respect to all 24 of Project 2061's instructional criteria (in part so that we can determine empirically their relative importance), we realize that we cannot do this fast enough for the results to inform later stages of our work. We are left with having to make logical predictions about which instructional criteria are the most powerful. We have also struggled with several issues related to how we select and train analysts to apply the criteria to the videotaped lessons.

Choosing a Subset of Criteria. We have used a number of approaches to narrow the list of criteria, including reviews of expert opinion and existing research and the application of various selection filters, such as which instructional criteria are most closely related to how students will be tested. For example, because the tests will ask students to apply their knowledge in new and unfamiliar situations and will use common misconceptions as distractors, perhaps the instructional criteria that focus on real-world contexts and misconceptions should be the basis for teacher observations. Based on these various considerations, five instructional criteria have come to the top of our list. These are criteria that address (1) effective representation of the mathematics, (2) the use of relevant real-world contexts, (3) teachers' use of probing and guiding questions in their interactions with students, (4) making connections among and between mathematical ideas, and (5) providing opportunities for students to practice the skills they are being taught.

The choice of which criteria to focus on first is important in planning for the second phase of our study, but ultimately the narrowing of the number of criteria is also important for scale-up purposes. The field needs to define a manageable set of teaching skills that they can concentrate on both in professional development and in other research studies of this kind. Our intention is to produce a refined list or a clustering of instructional criteria that can be used more broadly by educators. Our initial judgments about which are most salient will be tested empirically throughout the study.

Selecting and Training Analysts. As mentioned in Part I of this paper, each teacher is videotaped three to five times on lessons in which the target learning goal is being addressed. The lessons are then analyzed with respect to some number of the 24 criteria that we have used to define quality teaching. Analysts enter their observations into a Web-based utility that prompts them to respond to each criterion and to time code their observations. The decisions to videotape the lessons and to make such a precise analysis of teaching have had a number of implications for our study and for future research, as well as for the likelihood of large scale dissemination of professional development in the future.

Applying the instructional criteria with precision can provide deep insights into teachers' practice but is very labor intensive as well. A closer look at one criterion, "using representations effectively," demonstrates that level of precision. When examining a videotaped lesson, the analyst looks at the representations the teacher is using to convey the mathematical ideas to students. Representations may include equations, graphs,

diagrams, models, manipulatives, drawings, analogies, verbal statements, and simulations. The analyst identifies the key idea being taught and considers how closely the representation aligns with it, using the indicators discussed on page 7. Then the analyst examines how the teacher uses the representation with students. The analyst must determine the level at which the teaching and/or the activity meets each indicator. Each indicator is marked as "met," "partially met," or "not met." The analyst must also provide a detailed justification for each rating. The result includes a numerical rating for the criterion as well as a qualitative description of the effective use of the representation.

This is a time consuming process that requires analysts who have been trained in the use of both the criteria and the utility. It also requires them to have a deep understanding of the mathematics content so they can judge the alignment of a representation (or any other aspect of the pedagogy referenced in the criteria) with the content of the learning goal. A major challenge of the study has been how to select and train analysts who can accurately judge the quality of teaching that they observe on the videotapes. It should be noted that we already have considerable experience in training analysts from other work completed at Project 2061, both from our analysis of textbooks and our work in assessment. The analyses being done in this study, however, are somewhat more complicated in that they require that judgments be made about teaching in a dynamic, constantly changing classroom environment. Creating transcripts of each of the videotaped lessons would make the analysis considerably easier, but the cost of doing so for such a large number of videotapes is prohibitive. Regardless of how it is ultimately done, our success in training analysts will have a significant impact on the number of teachers we can involve in the second phase of the study.

(3) Professional Development. Designing and delivering professional development to teachers raise a number of additional questions and issues. These include reward structures for participation, questions of transfer and generalizability of the knowledge and skills, selection and use of videotapes, the extent to which professional development should be curriculum dependent, and the appropriate audience for professional development.

Reward Structures. A middle school teacher's day is filled with both academic and extra-curricular duties, with little time left for planning or for professional development. Research has shown that the most compelling reason for teachers to engage in professional development is the intrinsic reward they get from doing so. They must feel that the time they are spending is worthwhile. According to Elmore, "school personnel are more likely to work collaboratively to improve performance if the work itself is rewarding and if the external rewards support and reinforce work that is regarded as instrumental to increased quality and performance" (2002, p. 21). This places a great responsibility on us to provide teachers with something of value that will help them in their own teaching. In addition, extrinsic rewards are important for showing teachers that the time and effort they expend is valued. To meet the obligation of providing an experience that teachers consider to be both of value to them and valued by others, we must keep teachers' needs in the forefront at all times.

There are a number of approaches that we will take to ensure that the professional development we design takes teachers' needs into account. In the end, our professional development must fit within existing organizational structures including the time that is available to teachers. What we are trying to learn now is how each of these elements, including the consideration of reward structures and the need to work within existing professional development networks and time constraints, is related to successful professional development. Our efforts to date focus on the following:

- **Data-driven professional development.** Teachers are involved in active analysis of their curriculum, teaching, and student performance data and the possible relationships among them rather than being presented with information on general teaching methods. (See Falk, 2001; Smith, 2001; and Wilson & Berne, 1999, for discussions of data-driven professional development.) The use of their own textbooks and videotapes of their own teaching have enhanced teachers' engagement in the workshops.

- **Professional development schedule.** Because of the limited amount of time available during the school year, most of our professional development so far has taken place during the summer. Even then, one week is the maximum time that most teachers can spend. School-year workshops have been more difficult to organize, especially when there are few professional development days available and those are used by districts for their own specific activities. In addition, schools are reluctant, even with offers to pay for substitute teachers, to release teachers for a day. In one district we have been successful scheduling some sessions during district-wide professional development days, whereas in other districts we have used evenings. We are also experimenting with homework activities, but these place large burdens on teachers after having already spent a full day in school.

- **Teacher stipends.** Teachers are given a stipend for each day of professional development they attend. They also receive stipends for time spent on various assignments in which they develop lessons or prepare journals of their reactions to or reflections on their work. Each year, a few teachers in some districts are given support to attend the annual meeting of the National Council of Teachers of Mathematics.

- **Communities of practice.** We have not yet determined how to establish communities of practice (Cochran-Smith & Lytle, 1999; Wenger, 1998) either online or at the school and district level to support professional development. Nor have we determined the extent to which such communities of practice are logistically possible given the demands of the school day. We expect to ameliorate some of the structural problems by taking advantage of Web-based communication technologies. But we also recognize the importance of face-to-face contact and the need to establish actual (as opposed to just virtual) communities of learners. This will be a major challenge for us as we proceed through the study.

Transfer and Generalizability. Because professional development time is limited, we need to organize that time so that the principles learned can be generalized to new situations. We want teachers to be able to apply the ideas learned in the professional

development to other topics and in other classes that they teach. The questions we have to answer are what kinds of transfer are possible and what kinds of experiences over what periods of time are needed to get teachers to the point where such transfer takes place?

A key issue is how close the professional development experience should be to teachers' actual classroom experiences. When a teacher analyzes data from her own classroom on lessons she will teach again and again, transfer to other similar lessons and content is likely. If a teacher analyzes excerpts from another teacher's videotaped lesson using the same curriculum materials and teaching the same or similar content, will the teacher still grasp the principles being addressed? If so, will that knowledge transfer to her or his own teaching? What if the teacher analyzes video excerpts from teachers who are using an entirely different set of curriculum materials? In our study we will pay attention to issues of transfer by looking at the kinds of experiences that lead to transfer and the extent to which teachers are able to apply their new knowledge and skill.

A major consideration in this regard is the length of time it takes to achieve transfer. What is the best way to distribute that time? These choices affect decisions of breadth versus depth in the content of professional development and depend on what is possible for teachers to accomplish given limited time. In order for professional development to be taken to scale, it is necessary to consider the total time required to achieve the desired ends and how that time is distributed.

The Selection and Use of Videotapes. The use of classroom video is powerful and reveals how students think and how teachers interpret and react to students' ideas. Whereas using positive models from the classroom is helpful, we also want to show excerpts where classroom teaching is not well aligned with the learning goal or where it is not consistent with the instructional criteria. This raises questions of how much teachers are willing to reveal about themselves publicly, given that some of the most instructive episodes are ones where teaching does not match the instructional criteria or the content alignment expectations.

For example, a primary goal for the professional development in the summer of 2003 was to examine students' various representations of equivalence. We had videotapes of lessons from all four curriculum materials that addressed the number learning goal dealing with comparing equivalent forms of integers, fractions, decimals, and percents. In preparation for professional development, an intensive review of these lessons was undertaken to locate a range of instances where students used representations of equivalence in both standard and non-standard ways. In a videotaped lesson from the *Connected Mathematics* unit Bits & Pieces I, students are expected to move from using a fraction bar for modeling fractional parts to a mental part-part-whole analysis. A number of students used the fraction strip in new ways or developed alternative representations rather than move to the more abstract part-part-whole analysis. In one instance, the camera catches a small group interaction in which a student persists in explaining an alternative use of his fraction strips. The student attempts to make sense of the mathematics in his own terms. The episode was powerfully illustrative of both the challenges of the lesson and of the nature of student sense-making in the context of

shifting mathematical representations. However, it was also a case where the teacher may have felt as if she had been caught off-guard by the camera and might regard it as a blemish in an otherwise excellent performance.

To ensure that video excerpts are used in the most professional and productive way possible, we are developing a number of guidelines for their selection and use:

- Involve teachers in searching the videos for examples from their own classrooms. This scaffolds the creation of teacher knowledge and returns ownership of the excerpts to the teachers themselves.
- Provide a transcript of the video excerpts to be used in professional development. The interplay of visual record and written artifact allows for a deeper analysis and richer discussion than either alone. And when used alone, the transcripts provide a layer of anonymity not possible with a videotape.
- Shorter is often better. Much of the substance of an episode that is potentially educative for teachers happens within a small period of time in the give-and-take of actual instruction. This microanalysis may be central to impacting teacher understanding, skills, and attitudes. (Additional video may be used to provide a context for the analysis although context can often be achieved more easily and with less distraction through narrative description.)
- During professional development workshops where the videotapes are used, allow the teacher to introduce and set the context for her own video segment. A corollary is to allow the teacher whose practice is being made public (with her consent) the right of first commentary after the video excerpt has been viewed or transcript read. Although we did not use a formal protocol to accomplish this in our first summer institute, a more formal procedure for viewing video segments could be employed to good advantage.

Curriculum-dependent professional development. The professional development that we provide focuses on specific learning goals and on specific instructional criteria, and these learning goals and instructional criteria then get interpreted in terms of the materials that each teacher is using. The professional development is designed to be applicable to all four of the textbooks that the teachers are using. Discussions center on the instructional criteria and the learning goals, and the teachers work with cases from their own classrooms and textbooks. The professional development is offered separately in each state, but the format, content, and activities are the same. Only when analyzing the videotapes of their classes are the teachers paired with teachers using the same textbook so that the discussions become textbook specific.

Inevitably, this way of organizing professional development is going to lead to a certain amount of sharing of ideas and resources between teachers using different materials. Some of the texts are better aligned with the learning goals and some offer more pedagogical support for teachers than others. Sharing across text materials could very well lead teachers to use lessons and activities from the other texts that are not consistent with the design principles of the textbook they are using. Because we wanted, at least in the first phase of the study, to test the effect of the curriculum itself on teacher behavior and student learning, any amount of sharing of resources between teachers dilutes that

curricular effect. This is why we do not offer our own supplementary materials for use by the teachers.

There are other factors working against the idea of a "pure" application of the curriculum. For instance, we have found that teachers vary considerably in their commitment to their curriculum materials. Some teachers that we thought were using a given material later told us that they used it only as a supplement or, conversely, that they supplemented it heavily with other materials. Variation in the materials used and in the way those materials are used challenges our capacity to ground professional development in the practice of each teacher and to answer certain research questions. Whether it would be more effective to intentionally design professional development around the activities, contexts, and representations in specific texts or not is something that we will have to decide for the second phase of this study.

Audience. The teacher participants in our study vary according to the length of prior experience they have had with their respective mathematics curricula and the amount of professional development they have had on those materials. This means that in some cases we are working with teachers who have just begun to use the materials and may not yet have a full grasp of their intent. (See Ershler, 2001, for a discussion of the special challenges inherent in providing professional development to novice teachers.) In other cases the teachers are very familiar with the materials and are ready to move on to more in-depth analysis of their teaching of those materials. It is certainly of interest to know how familiar teachers need to be with the curriculum materials they are using before professional development like ours is effective, but because of limited time and resources, we are planning to limit the second phase of the study to relatively experienced teachers. It is important to point out that even though we will most likely not be able to do this ourselves, we do believe that testing the impact of this kind of professional development on novice teachers is a useful study. Similarly, it would be useful to know to what extent pre-service teacher education students can make use of these ideas and experiences.

(4) Student Assessment. Developing and refining appropriate instruments for measuring student learning within the context of this study are critical steps. To ensure the quality of the results of these assessments and their suitability for the purposes of our investigations, we have focused our attention on the following issues related to the assessment process.

Test Administration. An issue that we confronted in the second year of the study has to do with differing time intervals between administration of the pre-test and post-test for different classes. We had hoped to keep the interval constant for all classes by having a set time for administration of the tests. However, in the second year, some teachers were testing students on both the Data and Number learning goals or the Data and Algebra learning goals, and time constraints forced a number of teachers to modify the testing schedule. At the moment we do not intend to control statistically for time between test administrations, but we will note it as a possible explanation for results that appear to be irregular.

Scoring Accuracy and Efficiency. From production, to administration, to scoring of the tests, there are places where errors can occur. The challenge is to be mindful of the sources of error and to anticipate where and when they may occur. This is a particular problem when thousands of papers are being scored. For the first year of testing, students wrote their responses directly on their test papers, and the scores were then entered as numbers into a database. This was immensely labor intensive and expensive and presented many possibilities for error. In the second year, we used answer sheets that could be scanned optically. This reduced the labor but introduced other kinds of errors, such as "bubbles" not properly shaded in. We are now working on a Web-based version of scoring in which scorers will record student responses online. This method will also be tested for scoring the constructed response items. Our efforts are aimed at achieving maximum accuracy with minimum cost and expenditure of time on the part of scorers.

Assessment Maps. As noted earlier, the assessment maps provide the basis for the development of all test items and the distribution of items that are selected for the assessment instruments. The assessment maps result from an analysis and clarification of the meaning and intent of the learning goals. In Project 2061's previous work on science assessment, assessment maps were designed to provide instrument developers with a visual representation of the ideas related to a specific learning goal that might be tested. Those maps are derived from the Project 2061 conceptual strand maps (AAAS, 2001) that show the probable learning trajectories of ideas in science. Inherent in these learning trajectories is the notion of prerequisite knowledge. Assessment maps also include information about related ideas and students' misconceptions.

Assessment instrument developers can make use of the information contained in the assessment maps as appropriate. Developers can use information on misconceptions to directly test whether students hold those misconceptions or they can use that information to create plausible distractors. Similarly, instrument developers can use the information on prerequisite knowledge to diagnose learning difficulties by testing students' understanding of those prerequisite ideas. Maps also display ideas that come later in the learning trajectory, and those might be tested as well. The maps are not prescriptive but rather are aids for constructing tests, and the use to which they are put depends on the purpose of the test.

The maps that we have created for this study are a variation on the maps that were created for our earlier work with science assessments. They differ in that they do not identify a learning trajectory for the ideas that are part of the learning goal. Nor do they suggest which ideas from earlier grades might be needed in order to understand the ideas in the target learning goal. They map the terrain of relevant knowledge, but they do not make any claims about which ideas precede which others. Instead, they deconstruct the learning goal into its conceptual parts—the ideas and skills that are implied by the learning goal itself. The mathematics education experts that we consulted during the construction of the maps did not feel that it was possible to suggest a progression of knowledge for the learning goals. Therefore, each of the maps that we created for this study includes two or three sub-ideas, which in turn encompass an additional number of more detailed ideas and skills.

In this study we plan to compare the structure of our assessment maps with the empirical findings that come from the test results. Factor analysis will be used to determine if our hypothesized clustering of ideas matches the empirical data. We will also use students' responses on specific items to test whether certain learning trajectories do indeed exist. By combining theoretical and empirical perspectives, we will be able to refine the content of the specific assessment maps for each learning goal we are studying, and we should gain insights into how assessment maps should be structured more generally.

Whereas the use of specific learning goals as a focus of the study allows for an analysis of relationships at a precise level of detail, the assessment maps enable us to make those analyses at a still finer grain size. For each sub-idea, we will be able to link textbook treatment of that idea, teacher attention to that idea, and time spent in professional development on that idea to student outcomes (measured by our own instruments and by state assessment items that are specifically related to that idea).

A Work in Progress
In this paper we have laid out the design of a study that addresses the improvement of middle school mathematics teaching and learning through professional development. We have discussed a number of the issues and questions we have faced along the way. We bring this to the mathematics education community at this time for two reasons: (1) so that that we can inform others of the complexities of a study of this kind, and (2) so that we can receive feedback from you on the various issues that we have raised. We welcome any comments and suggestions that can help us to design professional development and seek answers to important research questions in the best and most effective ways possible. Please send your feedback to any of the authors of this paper.

References

Acquarelli, K., & Mumme, J. (1996). A renaissance in mathematics education reform. *Phi Delta Kappan, 77*, 478–484.

American Association for the Advancement of Science (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

American Association for the Advancement of Science (2000). *Middle grades mathematics textbooks: A benchmarks-based evaluation*. Washington, DC: Author.

American Association for the Advancement of Science. (2001). *Atlas of science literacy.* Washington, DC: Author.

American Association for the Advancement of Science (2004, July). *Project 2061. A procedure for the analysis and development of science and mathematics assessment tasks*. Manuscript in preparation.

Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is—or might be—the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher, 25*(6), 8–14.

Billstein, R., Williamson, J., Montoya, P., Lowery, J., Williams, D., Buck, M., Burkett, C., Churchill, L., Clouse, C., Denny, R., Derrick, W., Dolezal, S., Galarus, D., Kennedy, P., Lamphere, P., Merrill, N., Morse, S., Petit., M., Runkel, P., Sanders-Garrett, T., Seitz, R., Spence, B., Sowders, B., Tuckerman, C., Wenger, K., Wilkie, J., Wilson, C., & Winston, B. (1999). *Middle grades math thematics*. Evanston, IL: McDougal Littell.

Cochran-Smith, M., & Lytle, S. L. (1999). Relationships of knowledge and practice: Teacher learning in communities. In A. Iran-Nejad & C. D. Pearson (Eds.), *Review of research in education, 24*. Washington, DC: American Educational Research Association.

Cohen, D. K., McLaughlin, M. W., & Talbert, J. E. (Eds.) (1993). *Teaching for Understanding: Challenges for policy and practice*. San Francisco: Jossey-Bass.

Collins, W., Dritsas, L., Frey-Mason, P., Howard, A. C., McClain, K., Molina, D. D., Moore-Harris, B., Ott, J., Pelfrey, R. S., Price, J., Smith, B., & Wilson, P. S. (1998). *Mathematics applications and connections*. Columbus, OH: Glencoe/McGraw-Hill.

Collis, K., Romberg, T.A., & Jurdak, M. (1986). A technique for assessing mathematical problem-solving ability. *Journal for research in mathematics education, 17*(3), 206-221.

Elmore, R. (2002). *Bridging the gap between standards and achievement: The imperative for professional development in education.* Washington, DC: The Albert Shanker Institute. .

Ershler, A. (2001). The narrative as an experience text: Writing themselves back in. In A. Lieberman & L. Miller (Eds.), *Teachers caught in the action: Professional development that matters* (pp. 159-173). New York: Teachers College Press.

Falk, B. (2001). Professional learning through assessment. In A. Lieberman & L. Miller (Eds.), *Teachers caught in the action: Professional development that matters* (pp. 118-140). New York: Teachers College Press.

Fennema, E., & Franke, M. L. (1992). Teachers' knowledge and its impact. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 65–97). New York: Macmillan.

Hiebert, J., & Carpenter, T. (1992). Learning and teaching with understanding. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 65–97). New York: Macmillan.

Kesidou, S., & Roseman, J. (2002). How well do middle school science programs measure up? Findings from Project 2061's curriculum review study. *Journal of Research in Science Teaching*, *39*(6), 522-549.

Kulm, G., Capraro, R. M., & Capraro, M. M., & Hastings, E. (2002, April). *Increasing student achievement: Building on ideas and promoting thinking about mathematics.*

Lampert, M., & Ball, D. L. (1998). *Mathematics, teaching and multimedia: Investigations of real practice*. New York: Teachers College Press.

Lappan, G., Fey, J. T., Fitzgerald, W. M., Friel, S. N., & Phillips, E. P. (1998). *Connected mathematics*. Menlo Park, CA: Dale Seymour.

Lappan, G., Fey, J. T., Fitzgerald, W. M., Friel, S. N., & Phillips, E. P. (2000). *Connected mathematics*. Glenview, IL: Prentice Hall.

Romberg, T., Burrill, G., Fix, M., Middleton, J., Meyer, M., Pligge, M., Brendefur, J., Brinker, L., Browne, J., Burrill, J., Byrd., R., Christiansen, P., Clarke, B., Clarke, D., Cole, B., Dremock, F., Halevi., T., Milinkovic, J., Shafer., M., Shew., J., Schultz, K., Simon, A., Smith, M., Smith, S., Spence, M., & Steele, K. (1998). *Mathematics in context.* Encyclopedia Britannica Educational Corporation.

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

No Child Left Behind Act of 2001, 20 U.S.C. § 6301 *et seq.* (2002).

Shulman, L. (Ed.). (1992). *Case methods in teacher education*. New York: Teachers College Press.

Smith, M. S. (2001). *Practice-based professional development for teachers of mathematics*. Reston, VA: National Council of Teachers of Mathematics.

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: Free Press.

Wenger, E. (1998). *Communities of practice.* New York: Cambridge University Press.

Wilson, M. S., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of Research in Education*. Washington, DC: American Educational Research Association.

Zumwalt, K. (1998). Beginning professional teachers: The need for a curricular vision of teaching. In M. C. Reynolds (Ed.), *Knowledge base for the beginning teacher* (pp. 173-184), Oxford, UK: Pergamon.

Data: Diagram to Aid Assessment Task Design

learning goal

By the end of 8th grade, students should know that comparison of data from two groups should involve comparing both their middles and the spread around them. 9D/M4

Contexts for the "representations" and "summary measures" idea sets, the "key ideas" and the "learning goal":
* data presented in any form, e.g. raw, tabular, graphical
* calculators allowed

key ideas

The learning goal has two ideas embedded in it.

Students should **know that** comparing two sets of data must include both middle and spread. DK1

Students should **know how to** compare two sets of data using both middle and spread. DK2

summary measures

In order to compare middles and spreads, students need both procedural knowledge and conceptual understanding of the measures used to describe them.

representations

In order to compare sets of data, students need to know how to represent data in different forms and how to read and interpret various given representations.

Understand concepts of mean, median, and mode. Recognize the effect of additional data points on measures of central tendency. DS1

Determine the mean, median, mode, and range for a set, or sets, of data, presented in various forms. DS2

Zero is not a data point.

Select and use an appropriate measure of central tendency to describe a set of data, including mean, median, and mode. DS3

Describe the spread of a set of data, using measures such as range and quartile, and descriptors such as outliers, clusters ("clumps"), and gaps. DS4

Mean, median, and mode are the same.

Organize data into meaningful units. DR1

Select and construct an appropriate representation for a set of data to illustrate characteristics of central tendency or dispersions, such as boxplots, lineplots, and, stem & leaf plots. DR2

Read and interpret, various representations of a set, or sets, of data. DR3

Scaling errors.