

**DESIGN CONSIDERATIONS FOR EVALUATING THE EFFECTIVENESS OF
TECHNOLOGY-RELATED TEACHER PROFESSIONAL DEVELOPMENT**

Barbara Means, Robert Murphy, Hal Javitz, Geneva Haertel, and Yukie Toyama

Center for Technology in Learning

SRI International

D R A F T

August 14, 2004

(edited 10-20-2008)

DESIGN CONSIDERATIONS FOR EVALUATING THE EFFECTIVENESS OF TECHNOLOGY-RELATED TEACHER PROFESSIONAL DEVELOPMENT

Barbara Means, Robert Murphy, Hal Javitz, Geneva Haertel, and Yukie Toyama

Center for Technology in Learning

SRI International

INTRODUCTION

Federal support for the use of technology to enhance K-12 education has long recognized the critical role that teachers play in determining whether and how technology is used in the classroom. The Enhancing Education Through Technology (EETT) program, authorized as part of the No Child Left Behind Act of 2001, provides for formula grants to states to be used to improve student achievement through the use of educational technology. In recognition of the importance of providing teachers with appropriate training and support for this effort, the legislation stipulates that states spend at least 25 percent of their EETT funds on teacher professional development. Given the level of state grants under EETT (\$641 million in FY2003), the legislation thus supports *at a minimum* \$160 million in teacher professional development activities annually. While the legislation states that the EETT-funded professional development needs to be “ongoing,” “intensive” and “high quality,” it does not set forth specific criteria for meeting these requirements.¹

This paper seeks to explore potential evaluation designs suitable for an evaluation of technology-related teacher professional development by reviewing pertinent prior research, discussing key design issues and choices, and suggesting a recommended design and several variants of it that may inform future evaluation efforts.

¹ EETT guidance does say that high-quality professional development includes, but is not limited to, activities that: improve and increase teachers’ knowledge of academic subjects and enable teachers to become highly qualified; are an integral part of broad schoolwide and districtwide educational improvement plans; give teachers and principals the knowledge and skills to help students meet challenging state academic standards; improve classroom management skills; are sustained, intensive, and classroom-focused and are not one-day or short-term workshops; advance teacher understanding of effective instruction strategies that are based on scientifically based research; and are developed with extensive participation of teachers, principals, parents, and administrators.

One of the major challenges in executing this task is the lack of clearly codified, widely used “models” of technology-related teacher professional development (abbreviated as “TTPD” in the remainder of this paper). TTPD varies widely in terms of its goals, mechanisms, duration, and provider type.

Technology-related professional development may seek to do one or more of the following:

- Overcome teachers’ anxiety over or resistance to using computers
- Teach the mechanics of using specific pieces of hardware or software
- Teach, support, and stimulate adherence to a general pedagogical model in which students are active learners and technology use is integrated
- Train teachers in how to implement a specific curriculum unit and/or pedagogy that involves use of technology.

The training may use any number of delivery mechanisms, including:

- “Stand-up” face-to-face workshops or training sessions
- Computer-based instruction
- Web-based delivery

Some training efforts seek to put every teacher through the technology-related training, but a very common alternative is the “train-the-trainers” model, in which an initial set of individuals is trained and they are expected to provide professional development for the rest of the target group. Moreover, as will be discussed below, informal forms of professional development are regarded as effective supplements to (or even replacements for) more formal delivery mechanisms. These include activities such as self study, mentoring, study groups, and site-based assistance on demand.

The duration or intensity of professional development is another important dimension, and one that may be related to delivery mechanism. Teacher professional development is often criticized for expecting to change teaching practice dramatically through a single afternoon’s attendance at a lecture or workshop. As schools, districts, and states have explored options for extending the time engaged in professional development, they have emphasized longer summer workshops, regular schedules of weekly or monthly professional development sessions, self study (typically undertaken outside of the workplace and the

workday), and “study groups.” Formal courses are another option for increasing the number of hours devoted to professional development.

TTPD varies also in terms of the nature of the provider. While there are commercial providers (such as Classroom Connect and Knowledge Universe), much professional development is both designed and delivered by district staff. Institutions of higher education are another major provider. Teacher surveys suggest that colleagues, friends, and family members can be significant sources of technology learning also.

Finally, the extent to which the teachers participating in the professional development activities have chosen to do so may be an important variable: Some professional development is required by districts or states; in other cases, TTPD is a licensure or degree requirement. In still others, there are tangible inducements for participating in TTPD (such as receipt of a laptop computer or credit toward an increase in salary), or teachers may simply choose to participate because they are interested.

Given all the possible combinations of these five dimensions, it is clear that a rationale for narrowing the focus of an evaluation of TTPD to certain types of professional development activities is needed.

One reasonable basis on which to choose a focus is the frequency with which a particular type of TTPD is offered, on the assumption that the most commonly used forms are those on which most EETT funds are being spent. Until such time as data on the types of professional development being supported through EETT are available, we can turn to information available on earlier technology-related professional development.² The Professional Development Study conducted as part of the Integrated Studies of Educational Technology (ISET) included a 2001 survey of 1,273 elementary and secondary teachers drawn from a nationally representative sample of districts (Adelman et al., 2002). The survey asked teachers about both formal and informal forms of technology-related professional development and found that roughly the same proportion of teachers participated in both (76 and 78 percent, respectively). Among formal professional development activities related to technology, the within-district workshop or institute was the most common, with 66% of

² The ongoing National Educational Technology Trends Study (NETTS) will be conducting a survey of teachers that includes this topic in spring 2005 with a follow-up survey in spring 2007.

teacher respondents reporting they had attended one in the prior year.³ The dominance of this type of formal TTPD is apparent when compared to the second most frequently cited formal activity, workshops or conferences outside the district, which were experienced by just 21 percent of teachers. While no other form of formal TTPD occurred at a frequency comparable to that of the district workshop, three different forms of *informal* TTPD were experienced by just over half of the teachers in the survey sample: individual learning (reading journals, browsing the Internet for materials, etc.); going to Web sites to get educational technology information or materials; and informally working with peers, family, or friends on skills related to the use of technology in teaching.

The prevalence of a type of learning opportunity is not the same as its efficacy, of course. For each activity type in which they participated, teachers in the ISET survey sample were asked “To what extent did it prepare you to use educational technology in teaching?” Analysis of the survey responses showed that 60-65 percent of teachers answered either “a great extent” or “a moderate extent” for each of the four most common TTPD activities (district workshops, individual learning, Web sites, and informal work with peers, family, or friends). Thus, there were no differences among activity types in their effectiveness as perceived by participating teachers (Adelman et al., 2002).

The ISET survey provides an indication also of the content focus of teachers’ professional development in technology. More than half the teachers participating in a formal professional development activity received training on one or more particular software applications. The most commonly trained individual applications were productivity tools—email, word processing, Internet browsers, and desktop publishing and multimedia presentation software. This finding does not mean that TTPD is confined to promoting teachers’ computer literacy, however. Two-thirds of the teachers who had experienced formal TTPD said that it covered topics related to teaching in their primary content area. In addition, 59 percent said their formal TTPD covered how to “promote active learning” and 57 percent said that they were trained in the use of technology “to promote basic skills.”

³ The survey item wording was “within-district workshops or institutes focused on a specific topic, provided by or within the district.” There was no specification as to the nature of the topic.

Another criterion for selecting types of TTPD to evaluate would call for examining those models that have the strongest research base—those for which we have the strongest empirical evidence of effectiveness. To develop this perspective for the TTPD evaluation design, we undertook a review of the research literature, described below.

LITERATURE REVIEW

Background

To a large extent, the research literature on teacher professional development is organized around a search for features or elements of “best practice” rather than around studying the effectiveness of different models. Based largely on expert opinion, buttressed by survey and case study research, a conventional wisdom around desirable features of teacher professional development has arisen and been disseminated in the education literature (see, for example, Brand, 1997; CEO Forum, 1999; Little, 1993; Loucks-Horsley et al., 1998; Persky, 1990). Features generally cited as elements of best practice include:

- Extended duration (more than a single session)
- Direct relationship to the content the teacher teaches
- Coherence with teacher goals and standards for which teacher is responsible
- Appropriate to participants’ level of knowledge, skill, and interest
- Active learning opportunities
- Follow-up activities
- Participation of groups of teachers from the same school

Several survey-based studies (Adelman et al., 2002; U.S. Department of Education, 2000) have found that if teachers are asked to report on whether such features characterized the professional development they received and then are asked in other questions to rate the effectiveness of their TPD or to report on their use of practices encouraged by the training, there is indeed a correlation between the reported presence of these “best practice” features and teachers’ reports that the training was effective. Concern lingers, however, that certain teachers may implement the classroom practices of interest not because of any professional development they received but because of other aspects of their skills, preparation, or pedagogical philosophy or because of other supports within the classrooms, schools, and

districts where they teach that are correlated with the provision of “best practice” TTPD. It may be, for example, that high-performing schools and districts recruit good teachers, offer them support for classroom innovations, and offer professional development with the characteristics deemed “best practice.”

The U.S. Department of Education (2000) study of the Eisenhower Professional Development Program provides a stronger case than most of the large-scale studies by surveying teachers annually for three years about both the professional development they had received in the prior year and their teaching practices. In this way, the analysts were able to control statistically for differences among teachers in terms of their previous practice when investigating the relationship between professional development and teachers’ pedagogy. Although not focused on technology use exclusively, this study of math and science teaching included an examination of professional development and practice in what the researchers called “use of technology for higher-order learning”--defined as practices such as use of calculators or computers to develop models, use of calculators or computers for data collection, use of computers to write reports, and the use of computers to access the Internet. Study findings suggest that professional development on the use of a specific teaching strategy did in fact increase the use of that strategy; moreover, the effect was amplified when the features of “best practice” were present (e.g., professional development focused on the use of calculators and computers to develop models was more strongly associated with subsequent implementation of that practice than was the receipt of professional development on other types of content, and the use of the practice was even more likely if the professional development on the use of calculators and computers to develop models was conducted with a group of teachers from the same school).

Approach to the Literature Search

Within the resource constraints of this design paper, we conducted a modest search of the literature on technology-related professional development, with several goals in mind. First, we wanted to characterize the empirical studies examining the effects of technology-related professional development. Specifically, we sought to provide some indication of the extent to which they have examined effects on student learning, and the quality of the research on this topic. In addition, we hoped to be able to draw some generalizations about the effects of

TTPD based on the research literature in order to guide the development of important questions for future evaluations.

To represent the corpus of evaluation studies related to technology-related teacher professional development, we searched three databases: Educational Resources Information Center (ERIC), PsychInfo, and Dissertation Abstracts. All of these databases were accessed through FirstSearch (www.oclc.org/firstsearch). Our initial plan was to pull studies conducted in the last decade, and accordingly, we limited the ERIC and PsychInfo searches to articles published from 1994 to the present. Because of the much greater volume of dissertation research, we limited that search to the years from 1997 to the present. As part of our search criteria, we stipulated that studies that were focused on higher education or adult learning should not be included in the results. We also excluded studies related to assistive technology for special education.

The search output was then subjected to an initial screening for relevance. At this stage, 110 papers were removed because their main content was not an evaluation of technology-related professional development for K-12 teachers. (Examples of excluded papers include those focused on describing a program of professional development rather than measuring effects and papers dealing more broadly with the status of technology integration in schools.) This initial screening resulted in the identification of 89 papers as falling roughly within the scope of our review.

A second screening eliminated 50 additional papers that either covered pre-service rather than in-service teacher education, used teachers from outside the U.S., or that used technology as a means of training delivery but not as the object of the training or one of the desired outcomes (e.g., use of videoconferencing to provide professional development on how to teach an aquaculture course). The Appendix to this paper provides a more detailed description of the literature search process and the effects of applying various screens.

The resulting 39 study abstracts were heavily slanted toward dissertation studies (26), augmented by studies found in ERIC (11), and just 2 from PsychInfo. (Abstracts of the final 39 studies appear in the Appendix.) We coded the abstracts of the 39 empirical studies of technology-related teacher professional development for their data sources, outcomes, and design dimensions.

Findings from the Literature Search

Data Sources. Table 1 displays the types of data collected in the 39 studies of technology-related teacher professional development. As the table shows, teacher surveys are by far the most common type of data collection in TTPD evaluations (used in 26 of the 39 studies). Because a separate code was used for formal assessments of specific attitudes or characteristics (e.g., REALs Belief Inventory), the extent to which the TTPD evaluation research relies on teachers' paper-and-pencil reports of their own feelings, opinions, or behaviors is even greater than the numbers for teacher surveys indicates. Thirty-one studies used surveys, teacher assessments, or both as outcome measures. While less prevalent, outcome measures not dependent on teacher reports were used in a number of studies. Ten of the studies analyzed samples of teachers' work (for example, lesson plans) and 8 involved observations of classroom activities. Only 2 studies involved the collection of data from students (one survey and one assessment).

Outcomes. The majority of the studies examined multiple outcomes, according to their abstracts. The distribution of outcome measures used in these studies is shown in Table 2. The table suggests that the most frequently measured criterion used to evaluate TTPD is teacher use of technology, either in instruction or for personal use (n=22) followed closely by teacher-reported attitude toward technology use (n=18). Other commonly evaluated outcomes were teachers' (self-reported or assessed) technology-related skills (including the ability to craft lessons integrating technology) (n=8) and teacher ratings of the quality or effectiveness of their training (n=7). The first three of these outcome categories are areas which are likely to be the direct target of the TTPD, and hence should show positive change if the professional development is effective.

What is strikingly rare among these studies is the use of student learning as the outcome by which the effectiveness of TTPD is evaluated. Only 2 of the 39 studies whose abstracts were reviewed used a student learning outcome as their dependent measure (1 used students' gain in subject matter knowledge and 1 used students' self-reported technology proficiency). This state of affairs is problematic, given the emphasis on using EETT funds to support the use of TTPD with effects on student achievement expressed in Title II of NCLB. This search of the recently published TTPD research literature suggests that the empirical base that

would enable states to identify high-quality teacher professional development is woefully lacking.

At the same time, we feel it is only fair to discuss reasons why researchers would choose outcomes other than student achievement measures as the criterion for evaluating TTPD. Although enhancing student learning is the professed goal of nearly all teacher professional development, we note that the chain of events linking professional development experiences to students' achievement gains is both long and complex. Figure 1 provides a graphical description. (See also the model description in Lesgold, 2003). In order for technology-related professional development to have a positive effect on student learning, the TTPD itself must be effective in giving teachers the motivation, knowledge, and skills needed to implement classroom practices that will change the nature of student experiences in ways that produce enhanced learning. To the extent that any of these linkages is less than perfect, we can expect measured effects to get weaker as we move from left to right across the diagram in Figure 1. Moreover, even if the theory underlying the training is correct and the teacher training itself is exemplary in producing teachers with the motivation and skills to implement new and better practices, the teachers' ability to do so is likely to be affected by the characteristics of the students with whom they work and of the school environment within which they teach. Many a teacher has left a workshop fired up about the use of project-based learning or a new curriculum on higher-order skills only to be disillusioned the first time she tries to implement her new learning with a class of students with low-level basic skills or a propensity for disruptive behavior. When the trained innovation includes the use of technology, the teacher's inclination and ability to implement will be affected also by the level of the technology infrastructure and the technical support she has back in her classroom. If she can't get on the Internet when she needs to or the software crashes frequently, she is likely to reject practices that require these elements. Finally, in many cases, and especially those where teachers for different grade levels and content specialties are grouped together for general training in technology, teachers will not be able to apply what they learned in training directly to their own classrooms but rather will have to make adaptations and hopefully "transfer" their learning in appropriate ways. Given all of these factors that can affect both whether teachers implement the trained practices and whether the practice is effective in producing student learning, most evaluators have opted for testing a

more proximal outcome (e.g., teacher knowledge or the implementation of the target practices) in the logic model.

Designs. Finally, we examined the research abstracts in order to classify the studies in terms of a set of research design categories, adapted from categories used by Kennedy (1999) in a review of studies of pre-service education. Based upon a reading of the abstract, the first author categorized a study as:

- *Follow-up of participants* if teachers who had participated in some professional development experience were subsequently surveyed, interviewed, observed etc. These studies had no control groups; some did include multiple versions of a treatment but in these cases the data were purely qualitative, precluding statistical comparisons.
- *Longitudinal studies of change* if a structured instrument yielding quantitative scores was administered both prior and subsequent to a professional development experience.
- *Quasi-experimental comparisons* of participants and nonparticipants or of participants in different types of programs. (If pre- and post-tests were used with two or more groups, this category was assigned.)
- *Experiments* in which participants were assigned at random to the treatment professional development and a no-treatment control or a contrasting version of the TTPD.
- *Multiple regression* type analyses in which a large number of teachers with varying professional development experiences were surveyed and statistical techniques used to identify relationships between professional development experiences and teacher-reported attitudes, practices, or results with students related to technology use.

Table 3 shows the results of this classification exercise. (In contrast to the classifications shown in Tables 1 and 2, each study could receive only a single code, that judged to best fit its design.) Follow-up studies and quasi-experiments were the most common designs in this corpus (with 11 and 10 exemplars, respectively). Longitudinal studies and multiple regression approaches were each used in seven studies. Only three studies employed experimental designs, and none of these had a student learning measure as an outcome.

Substantive Conclusions

The studies obtained through our search of the literature were disappointing with respect to their ability to inform a rigorous evaluation design. Most of them suggested that

technology-related professional development is effective, but very few had measures of effects on students or designs that would rule out alternative explanations for observed differences. A number of studies compared on-line and face-to-face versions of TTPD (Minier, 2002; Sujo De Montes, 1999), generally finding them to produce equivalent outcomes.

Two of the three experimental studies in the corpus generated by our literature search had findings inconsistent with the best practices recommendations based on survey results.⁴ Heine (2002) conducted an experiment in which all teachers received professional development around technology but some were assigned to a condition in which cooperative learning approaches were used while others were assigned to complete their training on an individual basis. When the teachers' lesson plans incorporating technology were scored for quality, no differences were found between the two groups. In Strickland's (2000) study of a technology-related professional development program, teachers were assigned at random to a condition incorporating follow-up activities through a listserv or to a no-listserv condition. No differences between the groups were found when teachers were surveyed with respect to their Internet usage or confidence in their ability to use the Internet.

Limitations of the TTPD Research Literature

As noted above, our search found no recent published experimental studies measuring the effects of TTPD on student learning or achievement. The main body of empirical data concerning the effectiveness of TTPD consists of correlations between teacher-reported characteristics of the technology-related professional development they've experienced and teacher reports of their practices or attitudes with respect to technology. In addition to lacking credibility with respect to establishing a causal relationship between TTPD and student achievement, these research studies tell us very little about the content focus of the professional development. The survey-based recommendations for best practice appear reasonable (i.e., preferring longer training over short, professional development with follow-up during the year to the "one shot" approach, etc.), but they do not tell a state, district, or professional development provider anything about what the *substance* of the technology-

⁴ The third experimental study (Raupers, 1999) was a complex design in which alternate forms of teacher professional development, either matching or conflicting with a teacher's assessed learning style, were delivered.

related professional development should be. In proposing a design for the national evaluation of the effects of technology-related professional development, we will propose a multi-part study addressing this issue of substance. Before turning to the topic of that design, we briefly describe several studies that have influenced our thinking but that fell outside the parameters for our literature search.

Selected Research Findings

In a 1998 monograph for the National Science Foundation, Mary Kennedy reviewed studies of in-service professional development in mathematics and science. Kennedy noted that most of the literature on professional development was devoted to advocating certain matters of *form* (e.g., extended timeframe, participation of all the teachers in the school, opportunities for teacher collaboration) as opposed to aspects of content—a conclusion very similar to that we reached for the technology-related professional development literature. She then collected studies that used experimental or quasi-experimental designs to test the effects of teacher professional development programs on student achievement outcomes. Covering a timespan from 1983-1996, Kennedy located 12 studies of mathematics professional development and 7 for science. In her analysis, Kennedy attempts to disentangle form and content features, and she makes the argument that many of the so-called best practices with respect to the form of teacher professional development may really be attributable to differences in *content* associated with different form features. For our purposes, two of Kennedy's conclusions are noteworthy. First, Kennedy found that the effect sizes for science professional development programs were larger than those for mathematics—a difference she attributes to the fact that the science programs tended to focus on particular content which was also covered by the researcher-developed assessments used as the outcome measures. In contrast, the mathematics studies she reviewed tended to use general-purpose standardized tests (e.g., the Iowa Test of Basic Skills). Among the mathematics studies, there was considerable variation in the magnitude and even the direction of TPD effects. The largest positive effects were found for the professional development in Cognitively Guided Instruction (CGI) studied by Carpenter, Fennema, Peterson, Chiang, & Loef (1989). Kennedy interprets the effectiveness of that professional

development to the fact that it provided the most specific information about the mathematics content that should be taught and about how students learn that content.⁵

Kennedy sums up her meta-analysis and interpretation of the math and science teacher professional development literature in words we think are equally applicable to technology-related professional development:

In both mathematics and science teacher education, then, the content of the program makes a difference. Inservice teacher education programs that teach different content also differ in their eventual effect on student learning. Yet the content of inservice programs is rarely mentioned in discussions of how to improve the value of inservice teacher education. Instead, these discussions tend to focus on such issues as the total contact time spent with teachers, whether that time is concentrated or distributed, and so forth. (p. 12)

A very recent study undertaken by SRI's Center for Technology in Learning, provides an example of an experimental evaluation of the effectiveness of a content-oriented, technology-related professional development initiative (Roschelle, Tatar, Kaput, & Hopkins, in preparation). In the first phase of a larger study being conducted under the Interagency Educational Research Initiative, these investigators explored the effects of teacher professional development on the use of a SimCalc replacement unit and associated *MathWorlds* software (Kaput, 1992) on both teacher and student knowledge of the concepts of rate and proportionality. Twenty-three Texas seventh-grade teachers were assigned to either the experimental or the control group. Both sets of teachers received 16 hours of conventional professional development on these topics (through a math professional development program with a strong reputation in the state). In addition, teachers in the treatment condition received an additional 24 hours of professional development in the SimCalc approach to teaching rate and proportionality with support from the *MathWorlds* software. Content knowledge tests assessed both the ability to perform calculations and

⁵ We note that Cohen and Hill (2001) emphasized the analysis of mathematics content in their survey-based study of the professional development provided as part of California's mathematics reform initiative. These authors found that professional development activities that were both "substantial" and focused around specific curricula and assessments were more likely than other professional development to be associated with the desired changes in teacher practice (as reported on teacher surveys).

answer multiple-choice questions on rate and proportionality and the ability to express these mathematical concepts through symbols, graphs, and words. Though preliminary, the study found large differences favoring the treatment group, both for teacher knowledge and for student knowledge in the area of connecting concepts and expressing them through multiple representations. (Note that like the science professional development studies reviewed by Kennedy, this study used an assessment carefully matched to the content of the training.)

Given the need to examine the content of TTPD and the dearth of rigorous evaluations of the effects of professional development on student learning as background, we turn now to a consideration of design options for a national evaluation.

DESIGN ISSUES AND GENERAL APPROACH

As indicated in the discussion above, technology-related teacher professional development varies along many dimensions, and there are no widespread, highly specified TTPD models amenable to evaluation (i.e., no TTPD analogs to Success for All). Under these circumstances, framing a set of research questions with enough specificity to drive a research design is challenging. We believe that policymakers need to be involved in identifying priority questions for the evaluation (Means & Haertel, 2004), but provide some thinking and discussion here in hopes of informing their deliberations.

Selecting a Focus

Given the context of large-scale evaluation studies, it makes sense to focus on a research question that will help guide state and district policies and practices in ways that support NCLB. This suggests measuring effectiveness in terms of student learning outcomes.⁶ Moreover, our review of the research literature suggests that a large-scale study investigating the effects of TTPD on student learning outcomes would be a significant addition to the research base. But what kinds of TTPD should be studied and how tightly should the intervention's fidelity be controlled?

⁶ At the same time, it would seem appropriate to collect information about the "quality" of technology-related professional development being funded through EETT through survey items tapping the dimensions of "best practice" identified by Adelman et al. (2002) and the U.S. Department of Education (2000). This descriptive function could be undertaken as part of the NETTS teacher survey activity.

One of the implications we drew from reviewing the research literature is that much of TTPD is rather general in focus (promoting a pedagogy of active inquiry supported by technology or the use of general productivity tools). The kinds of software most commonly covered in TTPD, according to teachers in the ISET survey, are Internet browsers (74%), e-mail programs (66%), desktop publishing or presentation software (63%), and word processing programs (60%). In addition, 35% reported receiving TTPD on the use of technology to promote active learning. The experimental studies showing the strongest impacts on student learning outcomes, on the other hand, have been studies that looked at training teachers to use techniques and curriculum units related to specific content (arithmetic in the case of Carpenter et al., 1989; and precalculus concepts in the case of Roschelle et al., in preparation). It seems to us that it would be very useful for policymakers to have an evidentiary base for choosing to promote one or the other type of approach to TTPD. This suggests one comparison for the evaluation.

A third type of TTPD we think the evaluation might consider is one that is just emerging but that has been given additional impetus by NCLB—the use of technology tools to support data-driven decision making at the school level. In 2001, 36% of teachers completing the ISET survey said that they had had professional development around using technology to analyze student assessment results including state and district assessment data. Given the proliferation of technology products supporting this kind of activity since the passage of NCLB (Stringfield, Wayman, & Yakimowski, in press), it seems likely that professional development on how to use these systems to guide instructional practices is on the increase.

Thus, we would propose that a national TTPD evaluation consider measuring the effects of three classes of professional development:

- Broad-focus TTPD on technology-supported pedagogy and the use of productivity tools,
- Content-focused TTPD on the use of technology in teaching specific content or skills, and
- TTPD on the use of technology tools that support teachers' data-driven decision making.

Given the fact that most technology-related teacher professional development is locally designed and delivered, we would not propose selecting a specific TTPD offering for evaluation. Rather, we would suggest the development of a protocol for each class of TTPD, specifying essential elements (such as what constitutes a productivity tool for the purpose of the evaluation) and some minimum requirements for training duration (i.e., number of contact hours) but leaving the choice of how to implement the training (for example, through a summer institute, weekly work groups, or use of Web media) up to the provider.

Selecting Outcome Measures and Study Populations

We discuss outcome measures and study populations together because of their interdependence. Given the decision to measure effectiveness in terms of student academic learning, the selection of a particular instrument or set of instruments for assessing that learning is still a complex issue. We can influence the apparent effectiveness of one form or another of TTPD through our selection of the outcome measure. From a policy perspective, the standardized tests used in accountability systems are the most pertinent measures of student learning. In addition, nationally normed standardized tests have the reliability needed for measuring gains from one point in time to another. On the other hand, these assessments are widely criticized on the grounds that they are not sensitive to instruction and are highly correlated with student and school socioeconomic variables. Because content-focused TTPD involves specific subject matter by its definition, it is likely to deal with the content tapped by only a small number of items on a standardized test that covers an entire grade-level curriculum. The most prudent course would appear to be the use of multiple measures—a content-specific test (e.g., an assessment of skills with decimals and fractions) and a broader assessment of the type used in state accountability systems (e.g., fifth-grade math).

This decision has implications for the specification of the treatment and the study population. It suggests selecting a subject domain in which there are both widely used curriculum units supported by technology and high-quality tests of content knowledge and skill. Potential candidates would include writing, algebra, geometry, physics, and biology.

While the broad-focus TTPD and TTPD on data-driven decision making would be applicable to teachers in any academic subject and grade level, we would lose the ability to

compare relative effect sizes across TTPD types if we confound type of professional development with grade and outcome measure. For this reason, we would recommend that all three experimental treatments be evaluated with teachers and classes in the same grade and content area. (Adding additional grades and content areas to the study of all three treatments would increase the generalizability of the evaluation findings, but would also increase study size and costs.)

Design Options

Given the goal of ascertaining effects on student academic learning, the recommended design is a random-assignment experiment. Only by recruiting a sample of willing participants and then assigning them to treatment or control conditions at random can we make a strong case that it was in fact the TTPD treatment and not some pre-existing differences in teacher, school, community, or student characteristics that produced any observed difference in academic performance.

Nevertheless, we recognize that random-assignment experiments are not widely used in educational research. The essential difficulty is not one of design or analysis but rather one of effective recruitment of participants and assurance of fidelity to the assigned condition (Shadish, Cook, & Campbell, 2002). These concerns often lead evaluators to opt for quasi-experimental designs and the attempt to use statistical controls to rule out competing explanations of observed effects.

Major cost elements of an experimental evaluation—measuring the fidelity of treatment implementation and assessing student learning gains—would be the same whether an experiment or a quasi-experiment is implemented. The decision boils down to weighing the greater persuasiveness of an experiment for the research and policy communities against the additional costs of participant recruiting. We believe that a carefully thought through approach to recruiting can minimize the additional costs.

The recruiting process differs depending on whether random assignment is planned for the student, teacher, school, or district level. Given that the treatment is applied to teachers, random assignment of students within classes is not sensible. Random assignment at the teacher level has benefits in terms of obtaining statistical power with a smaller sample of schools—a benefit which must be weighed against the potential for contamination (cross-

over) between treatment and control classes in the same school. The alternative would be to recruit districts and then randomly assign schools within districts, but this raises questions about potential “school effects” when a small number of schools volunteer within a participating district (i.e., differences between post-test scores of treatment and control schools in a district could be attributable to pre-existing differences among the schools).

With either teacher- or school-level random assignment, some districts are likely to be bothered by the fact that not all teachers (or schools) receive the treatment. A delayed treatment design, in which every teacher (or school) in the sample eventually receives the treatment has provided a workable solution to this issue for other studies (Slavin, 2002). Once the sample of teachers (or schools) is selected, half are assigned to receive the treatment in Year 1 and half in Year 2. The design thus permits not only a comparison of the effects of receiving the treatment versus no treatment in Year 1 of the study but also a comparison of experimental teachers’ second year of implementation versus their first after Year 2. (The delayed treatment group’s outcomes after treatment can be compared to those before treatment as well.)

Issues of Interpretation

A potential limitation on the interpretability of the findings of this kind of study stems from the fact that even the most effective teacher professional development can yield effects on student learning only if (1) the training provides the teacher with both the motivation and the knowledge and skills to implement the recommended practices, (2) the context within which the teacher works meets minimum requirements for implementing the practices encouraged by the training (e.g., the necessary software, technical support, and social supports are present) and (3) those practices result in student experiences which in fact enhance learning. In a sense then, a test of the effectiveness of the TTPD is a simultaneous test of the effectiveness of the training *and* the adequacy of the schools’ physical and social infrastructures *and* of the assumption that the encouraged teacher practices promote student learning. If we see positive effects of the training, we can assume that all of these prerequisites were satisfied. If we do not see any effects of the training, however, we will not necessarily know the locus of the problem. It will be important for the evaluation design to

incorporate components examining context and implementation to support accurate interpretation of findings.

Some problems can be avoided by stipulating requirements for the study sample that insure that schools provide a context within which the trained uses of technology *could* be implemented (i.e., the technology is available and use is encouraged and supported). At the same time, the study needs to keep in mind the focus on Title I eligible populations, which are often served in schools that have less Internet access (in terms of the student:computer with Internet ratio) and fewer full-time technology support staff than do schools serving more affluent students (NCES, 2003).

Another issue for the study is understanding (and documenting) the nature of the control condition. In many experiments, a treatment condition is contrasted with a no-treatment control. This evaluation will not be able to implement this sharp a contrast. It is reasonable for the evaluation to try to make sure that teachers in the control condition do not receive the particular EETT-funded professional development that is the object of the study. It would not be feasible, however, to rule out all forms and sources of technology-related professional development for the study's duration. Researchers could not prevent teachers from enrolling in college courses, finding Internet resources, or learning from family and friends. Realistically, the control condition will represent the "ambient level" of technology-related professional development teachers receive without special federal programmatic support rather than the total absence of opportunities to learn about technology integration.

DESIGN FOR AN EVALUATION OF TECHNOLOGY-RELATED PROFESSIONAL DEVELOPMENT

Research Questions

As noted above, we suggest that the evaluation of technology-related professional development focus on specific types of TTPD that are deemed to be both widespread and likely to be supported with EETT funds and that have policy relevance under NCLB. We have identified three types of TTPD that we judge to be strong on one or both of these criteria: broad-focus TTPD on technology-supported pedagogy and the use of general productivity tools; content-focused TTPD on the use of technology-supported curriculum

materials in an area of science or mathematics; and training on the use of technology tools to support teachers' data-driven decision making about instruction.

For each of these types, the primary research question, around which the experiment is designed, would be:

1. What effect does this type of technology-related professional development have on student academic achievement?

Corollary questions would be:

2. To what extent does this type of technology-related professional development lead teachers to implement the practices that are the targets of the training?
3. What contextual features are associated with larger or smaller effects on implementation of the target practices?
4. How strongly is implementation of the target practices related to positive student outcomes?

A decision would need to be made as to whether to structure the study as one large experiment comparing three different treatment groups or as an aggregation of three experiments, one conducted to test the effects of each of the three classes of TTPD. Our thinking has tended to follow the latter strategy, in large part because we expect school districts to be the primary designers and developers of TTPD, and we think that districts are unlikely to be willing and able to mount an arbitrarily assigned type of TTPD, yet alone three different types. If districts volunteer to be in a study of a particular type of TTPD, on the other hand, it is then feasible to stipulate that not all teachers or schools within the district receive the treatment in the first year and that random assignment is used to determine who gets the delayed treatment. This requires the presence of a control condition within each district—and in effect results in three separate experiments on the effects of three classes of TTPD.

Methodology

Level at Which to Implement Random Assignment. The required sample size depends upon the level at which the random assignment is executed (district, school, teacher/classroom, or student). The lower the level at which the random assignment is done (i.e., the closer to the student level), the fewer the number of schools that would have to be recruited and assessed in order to detect an effect of a given magnitude.

As noted above, the two most feasible alternatives for an experimental evaluation of TTPD are random assignment at the school level and assignment at the teacher/classroom level. The latter would require a smaller sample of schools, and hence would be more cost effective. Balanced against this consideration would be concern about the study's ability to maintain a distinction between treatment and control classrooms, as some teachers observe or hear about uses of technology made by their colleagues teaching classes in the same school. This concern appears particularly relevant in cases where the technology dealt with in the TTPD consists of web-based resources or tools.

On balance, we decided to assume teacher-level random assignment for purposes of laying out a design and estimating costs. As we thought about the threat of "contamination" across conditions within schools, it prompted a clarification of the intervention under test. As argued above, a condition that is literally a "no treatment" control is an impossibility. We can not prevent teachers from having access to any technology or deny them all support for using technology tools. What the evaluation would test is the effect of adding to the "ambient level" of technology access and support through the explicit design and delivery of federally supported professional development designed to motivate and equip teachers to make good use of technology in their practice. As long as teachers in the control condition do not participate in these formal professional development activities and are not required to use the technologies that are covered in those training offerings, we can test our research question about the effects of professional development.

Sample Sizes. Computing the required sample size requires knowledge of the size of effect the study wants to be able to detect as well as the level at which random assignment is conducted. Given the logic chain connecting TTPD to effects on student achievement, we would expect any effects of the training on teachers' motivation and practices to be

attenuated in some schools by features of the context in which the teacher works as well as by whatever limits are present in the potency of the trained practices.

With only a single experimental study in our literature review relating TTPD to student achievement, we have little empirical basis on which to base our assumption about an effect size. A somewhat broader sample of experimental studies is available in Kennedy's (1998) meta-analysis of the effects of teacher professional development in mathematics and science. Kennedy found an average standardized effect size of .15 for basic math skills and .28 for math reasoning and problem solving skills (both averages of 17 effects). A stronger effect size of .42 was reported in the studies of science teacher professional development examined by Kennedy (average of 13 effects).

Based upon the description of selected studies above and a sensitivity towards finding effects on student achievement that will be meaningful to policy makers, we would recommend that the study be designed with power sufficient to detect an effect size of .25.

The required sample size depends also on the "clustering effect" or the amount of variation in the outcome measure that can be explained by the variation between the units, or clusters, that will be subjected to random assignment. The "clustering effect" arises because the outcomes of individuals within clusters are likely to be more similar than those of individuals across clusters. (The statistical measure of the "clustering effect" is commonly referred to as the "intracluster" or "intraclass" correlation coefficient.)

Using estimates of intracluster correlation for reading and mathematics developed by Mathematica as part of the design study for the Evaluating Educational Technology Interventions (EETI) study (Agodini, Dynarski, Honey, & Levin, 2003) and a model for calculating needed sample sizes developed by Raudenbush and Liu (2001), we estimate that with random assignment at the teacher level, each TTPD study would require participation of 45-57 classrooms, with the higher end of the range required for mathematics outcomes. (In the elementary school achievement test data analyzed by Agodini et al., cluster correlations are higher in mathematics than in reading, requiring the larger number of classrooms for experiments with mathematics outcomes). If it were deemed important from a policy

standpoint to be able to detect smaller effects, larger samples of classes would be needed.⁷ The EETI study, which has similar recruiting requirements to the proposed TTPD evaluation, is getting an average slightly in excess of 3 teacher volunteers per participating school. Using that estimate, a power requirement for 57 teachers would call for around 20 schools.⁸ Similarly, based on the EETI experience, a study could expect to obtain an average of about 3 schools per participating district, suggesting the need to recruit 7 or so districts for participation in each of the three TTPD experiments.

Recruiting and Incentives. A strategy for recruiting study participants has been alluded to above is described here in more detail. We would suggest that evaluators seek the cooperation of the EETT program office in recruiting states to participate in the TTPD evaluation. Hence, the first step would be to obtain a sample of states willing to work with the evaluation team and providing a reasonable representation of state activity levels with respect to supporting educational uses of technology. Participating states would be asked to structure their subgranting process to be compatible with the recruitment of districts for one or more of the TTPD evaluation studies. In these states, districts applying for EETT subgrants would have the option of indicating a willingness to participate in one of the three TTPD experiments. At that time, information pertaining to the district's capacity for supporting the relevant treatment condition (e.g., number of schools serving the selected grade level(s), technology infrastructure, availability and experience of teacher trainers, etc.) could be collected. Ideally, districts participating in the experiment would receive some "enhancement" in their EETT subgrant to incent and cover the costs of a high-quality provision of training. Ideally, this funding would be part of the state EETT grant rather than an element of the research budget. In addition, it would be helpful for the evaluation budget to include a pool of funds to cover equipment upgrades or other technology costs associated with providing the TTPD treatment.⁹

⁷ For example, to detect an effect size of .20, it would take approximately 70 classrooms for reading achievement and 90 classrooms for mathematics achievement.

⁸ In contrast, the alternative of random assignment at the school level would require roughly 57 schools to detect an effect of the same magnitude. We considered this alternative to be unacceptable from a cost standpoint.

⁹ The availability of such a pool of funds for technology support has been a powerful selling point for the Evaluating the Effectiveness of Technology Interventions (EETI) study.

Once participating schools within districts are identified, we recommend the designation of a study point of contact for each school and an annual payment of \$300 to compensate that individual for assistance in arranging surveys and student assessments. We assume that support for teachers' participation in the professional development would be paid from the district's EETT subgrant. The evaluation budget should include some payment for both treatment and control teachers (say \$50 a year) as an incentive for returning teacher surveys in order to obtain a high response rate.

Data Collection. The major types of data collection entailed by this design are discussed briefly below:

- *Information on school demographics, technology infrastructure, and other contextual variables.* Random-assignment of teachers within school to immediate or delayed treatment conditions should ensure that the two groups of teachers are teaching in comparable contexts. Nevertheless, it is still prudent to check on variables such as teacher's education and experience levels. It is also necessary to document that each participating school had at least the minimum level of technology infrastructure and supports needed to implement the practices that are the targets of the TTPD under study. To the extent that participating schools vary in how much more than the minimum they offer, the evaluation can explore relationships between context variables and both implementation practices and student outcomes (Research Question #3).
- *Pretest scores on the target student learning outcomes.* Because students start at different places, even in classes within the same school, the evaluation will want to be able to control for prior achievement in its analyses. Any content-specific test is likely to have to be administered by the study; it may be possible to use scores from the state's accountability test as the broad pretest achievement measure, given appropriate measures to protect the confidentiality of individual student scores.
- *Document and observational data on the nature of the TTPD provided.* The evaluation will need to perform a "manipulation check" ensuring that the type of TTPD provided to the immediate treatment group does indeed fulfill the

requirements for the class of professional development under study. A combination of review of documents developed and used by the training providers and observation of a sample of training sessions can fulfill this requirement.

- *Measures of classroom implementation of the target teacher practices.* The TTPD will encourage teachers to implement specific practices—the integration of technology tools into student investigations or projects, the implementation of specific technology-supported curriculum units, or the use of technology supported achievement data analysis to make instructional decisions. The most cost-effective way to capture implementation levels would be through a teacher survey. All teachers in the evaluation (immediate and delayed treatment in each of the three experiments) should be asked about all three sets of practices.

Survey data may be supplemented with some level of classroom observation, but we are not recommending a major investment in such observations. The major difficulty is the fact that the target teacher practices are not things that teachers do all day, every day throughout the school year. One either has to conduct frequent, expensive site visits to have a reasonably high probability of observing the practices or to make explicit advance arrangements to come on a day when they will be executed—in which case they are “staged” and not necessarily representative of normal practice.

- *Post-test scores on the target student learning outcomes.* At the end of the school year, students in the participating classes would receive both the specific content assessment and a general achievement test. Because states and districts use different tests for accountability purposes and we want to be able to pool data across sites, we recommend that the evaluation obtain and administer the post-test instruments.

Analysis. We have envisioned the evaluation as a cluster of three experiments, each focusing on a particular type of TTPD. Analysis for the main research question (Research Question #1) in each experiment follows a relatively straightforward logic: an effect size is computed as the difference between the mean student achievement score for the treatment

classes minus that for control classes, divided by the standard deviation of the control classes. Before the simple comparison is made, however, multi-level regression modeling would be used to refine the estimates of student scores to take into account both prior achievement (measured in the pretest) and to adjust for context variables that covary with the outcome measure. The use of a multi-level or hierarchical regression approach will also account for any “clustering effect” that is present and provide more accurate estimates of the precision of the effect sizes calculated.

A similar analytic approach could be used in the analysis of classroom implementation of the practices that are the explicit target of the professional development variable (Research Question #2). However, because this outcome variable would be derived from scales constructed from survey item responses (one teacher per class), rather than from student assessments (an average of 20-25 per class), such an analysis would require data from more teachers. Mitigating this requirement to some extent is the fact that one would expect a much larger effect size for the teacher practices that are the target of the TTPD. After a review of studies using teacher practices as the outcome measure, a separate power calculation would need to be conducted to ascertain the number of schools required to evaluate this effect. Given the focus of the proposed evaluation on TTPD effects on students, we would propose an alternative analytic approach for Question #2 rather than any major expansion of the sample.

To address Research Questions #3, we would suggest an analysis of covariance, testing the ability of treatment variables (perhaps a 3- or 4-point treatment scale, reflecting the fact that some of the training will have been “better” in terms of intensity and incorporation of features considered best practice) and a few key contextual variables (e.g., school demographics, technology infrastructure) to predict teachers’ implementation of the classroom practices that are the target of the training. Finally, the evaluation could address Research Question #4 through hierarchical regression modeling, using the measures of implementation of the target practices as predictors of student achievement scores (with student pretest scores and treatment condition as covariates).

The analyses described above would be applied to each of the three experiments incorporated into the evaluation. Rather than comparing outcomes of the three types of

professional development to each other statistically, we would propose a simple summary and qualitative comparison of effects. Table 4 provides a shell for a table providing the kind of summary we envision. The evaluation's primary contributions to policymakers would be evidence of the effectiveness of each class of professional development and information about the implementation and contextual features related to positive outcomes for that class. Secondly, there would be some evidence of the relative strength of the effects produced by the three different classes of professional development as well as an indication of each approach's relative robustness—i.e., their ability to produce significant effects across a wide range of implementation and contextual variables (i.e., some types of TTPD may produce positive effects only in very supportive contexts).

Advantages, Disadvantages, and Costs

An evaluation of this complexity has many cost components, and costs can vary widely depending on decisions such as the number of types of TTPD to study and the selection of a grade and content focus and on decisions concerning the minimum effect size to be detected. Without going through a detailed costing exercise, we can rely on comparisons to other studies with which we're familiar to provide a rough sense of level of effort. A national evaluation of this scale would require the better part of a year for design and instrument development activities plus OMB clearance. A summer and full school year would be needed for the provision of professional development plus the collection of data from teachers and students. If the decision is made to collect data also on the second implementation year for the immediate treatment group and the first year after TTPD for the delayed treatment group, a second year of training and implementation would be required. In any event, data analysis and report preparation and review processes are likely to require a year subsequent to the end of data collection. Depending on the relationship of the start date to the academic year, the study could be expected to require 36-48 months.

Below, we provide some "guesstimates" as to the level of resources the design we have sketched out would require. If the three experiments were run concurrently as we've proposed, there should be economies in design and recruiting. A very rough cost estimate for running all three experiments and collecting and analyzing one year of data for those classes whose teachers did and did not receive the TTPD would be something in the

neighborhood of \$7-8 million.¹⁰ Cutting back to two types of TTPD would reduce costs, likely to something in the neighborhood of \$5.5-6 million. If the decision were made to include a second year of data collection and analysis so that teacher's second year of implementation could be compared to their first and a "replication" could be obtained by comparing post-training outcomes for the delayed treatment group to their pre-training outcomes, we estimate that costs for evaluating three types of TTPD would rise to about \$9-10 million.¹¹ Detecting smaller effects or including additional content areas or more than one grade level in the study would increase costs.

To put these cost estimates into perspective, it is important to remember that given the level of TTPD funded under EETT, over \$600 million in federal funds would be spent on technology-related teacher professional development during the four years of the evaluation study. Given that level of investment, a \$9 million study, as proposed here, does not seem unreasonable.

CONCLUSION

In considering recommendations for a national evaluation of the effectiveness of technology-related teacher professional development (TTPD), we have proposed a set of three concurrent experiments, each examining effects for an important class of TTPD. Our choice of this evaluation design was driven by policymakers' need to provide guidance to states and districts on the basis of evidence of TTPD effects on student learning. In our opinion, no other design would make a compelling case for a causal connection between TTPD experiences and outcomes for the teachers' students. A quasi-experiment in which districts or teachers volunteer for the technology-related professional development would always be susceptible to arguments that the volunteers differ from non-volunteers in important ways, no matter how many statistical adjustments are made. Moreover, we have argued that a rigorous quasi-experimental design would be nearly as expensive to implement

¹⁰ It should be noted that none of these figures is based on a cost analysis, which would itself be a labor-intensive activity. Such an analysis should be conducted at the point when some basic decisions about the scope of work for the evaluation have been made.

¹¹ Given prevailing wisdom that teachers need several years of experience integrating technology into their practice to do so effectively, there would be a good argument for including a second year of data collection. Even if data are not collected in the second year, however, the professional development could be supplied to the delayed treatment teachers under their districts EETT subgrants.

as a true experiment, yet would fail to provide as compelling a case with respect to effects on student learning.

The design we have proposed would not answer all of the interesting questions about technology-related professional development. There are still many things we would like to know concerning teachers' motivation for selecting to engage in professional development activities, the content focus of their TTPD, how much of their TTPD covers topics related to teaching in the core areas of reading and math, the administrative support teachers have for integrating technology into their instruction, and so on. We believe that descriptive data concerning the characteristics of the professional development provided through EETT can be captured through survey and case study approaches—and in fact, the ongoing NETTS project will provide this kind of data. The unique contribution of the experiments we propose here is the test of the hypothesis that the TTPD funded under EETT produces measurable improvements in student achievement. Evaluating whether EETT has met the goals expressed in its authorizing legislation requires this kind of evidence.

REFERENCES

- Adelman, N., Donnelly, M. B., Dove, T., Tiffany-Morales, J., Wayne, A., & Zucker, A. (2002). *The Integrated Studies of Educational Technology: Professional Development and Teachers' Use of Technology*. Menlo Park, CA: SRI International.
- Agodini, R., Dynarski, M., Honey, M., & Levin, D. (2003). *The Effectiveness of Educational Technology: Issues and Recommendations for the National Study*. Princeton, NJ: Mathematica.
- Brand, G. A. (1997). What research says: Training teachers for using technology. *Journal of Staff Development*, 19, 1. Available at <http://www.nsd.org/library/publications/jsd/brand191.cfm>
- Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C. P., & Loaf, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American Educational Research Journal*, 26, 499-531.
- CEO Forum. (1999). *Professional Development: A Link to Better Learning*. Author.
- Cohen, D. K., & Hill, H. C. (2001). *Learning Policy: When State Education Reform Works*. New Haven: Yale University Press.
- Heine, J. M. (2002). Staff development methods for planning lessons with integrated technology. *Dissertation Abstracts International* 63(12A): 72.
- Kaput, J. (1992). Technology and mathematics education. In D. Grouws (Ed.), *A Handbook of Research on Mathematics Teaching and Learning*. New York: MacMillan, pp. 515-56.
- Kennedy, M. M. (1999). The problem of evidence in teacher education. In R. A. Roth (Ed.), *The Role of the University in the Preparation of Teachers*. London: Falmer Press, pp. 87-107.
- Kennedy, M. (1998). *Form and Substance in Inservice Teacher Education*. Research Monograph No. 13. Madison, WI: University of Wisconsin, National Institute for Science Education.

- Lesgold, A. (2003). Detecting technology's effects in complex school environments. In G. Haertel & B. Means (Eds.), *Evaluating Educational Technology*. New York: Teachers College Press, pp. 38-74.
- Little, J. W. (1993). Teachers' professional development in a climate of education reform. *Educational Evaluation and Policy Analysis*, 15(2), 129-151.
- Loucks-Horsley, S., Hewson, P., Love, N., & Stiles, K. (1998). *Designing Professional Development for Teachers of Science and Mathematics*. Thousand Oaks, CA: Corwin.
- Means, B., & Haertel, G. D. (2004). *Using Technology Evaluation to Enhance Student Learning*. New York: Teachers College Press.
- Minier, R. W. (2002). An investigation of student learning and attitudes when instructed via distance in the selection and use of K--12 classroom technology. *Dissertation Abstracts International* 63(03A): 81.
- Persky, S. E. (1990). What contributes to teacher development in technology? *Educational Technology*, 30(4), 34-38.
- Raudenbush, S.W., & Liu, Xiao-Feng (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods*, 6(4), 387-401.
- Raupers, P. M. (1999). Effects of Accommodating Perceptual Learning Style Preferences on Long-Term Retention and Attitudes toward Technology of Elementary and Secondary Teachers in Professional Development Training (Elementary Teachers). *Dissertation Abstracts International* 59(11A): 129.
- Roschelle, J., Tatar, D., Kaput, J., & Hopkins, W. (in preparation). Scaling Up Opportunities to Learn More Complex and Conceptually Difficult Middle School Mathematics.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.
- Slavin, R. E. (2002). Evidence-based education policies: Transforming educational practice and research. *Educational Researcher*, 31(7), 15-21.

- Strickland, J. S. (2000). An investigation of the effects of using a LISTSERV discussion group after traditional workshops to sustain staff development: Training K--12 teachers to use Internet resources. *Dissertation Abstracts International* 61(09A): 126.
- Stringfield, S., Waymon, J. C., & Yakimowski, M. (in press). "Scaling up" data use in classrooms, schools, and districts. To appear in C. Dede, J. Honan, & L. Peters (Eds.), *Scaling Up Success: Lessons Learned from Technology-based Educational Innovation*. San Francisco: Jossey-Bass.
- Sujo De Montes, L. E. (1999). The Use of Internet-Based University Courses as a Tool for Professional Development of K-12 Teachers (Distance Education, Inservice). *Dissertation Abstracts International* 60(06A): 216.
- U.S. Department of Education. Office of the Under Secretary, Planning and Evaluation Service, Elementary and Secondary education Division. (2000). *Does Professional Development Change Teaching Practice? Results from a Three-Year Study*. Washington, D.C.

Table 1
Data Sources Used in Studies of Technology-related Professional Development

	Total	Dissertations	ERIC	PsychInfo
Teacher Data				
<i>Survey</i>	26	18	7	1
<i>Assessment</i>	11	8	1	2
<i>Work Sample</i>	10	7	2	1
<i>Interview</i>	9	6	3	0
<i>Observation</i>	8	4	3	1
<i>Focus group</i>	1	0	0	1
Student Data				
<i>Assessment</i>	1	0	1	1
<i>Survey</i>	1	0	1	0
Document Review	1	0	1	0
Not Specified	3	1	2	0

Note: Multiple data sources could be used by a single study.

Table 2
Outcome Variables Used in Studies of Technology-related Professional Development

	Total	Dissertations	ERIC	PsychInfo
Teacher Tech Use	22	15	6	1
Teacher Tech Attitude (comfort, phobia, anxiety, etc.)	18	13	4	1
Teacher Tech Skills (use software, write lesson plans with tech, etc.)	8	5	3	0
Perceived Training Quality	7	6	1	0
Student Tech Use/Skill	4	1	3	0
Teacher Knowledge	3	2	0	1
Perceived Benefits for Students	2	2	0	0
Other or Not Specified	3	1	1	1

Note: Multiple outcome variables could be used by a single study.

Table 3**Designs Used in Studies of Technology-related Professional Development**

	Total	Dissertations	ERIC	PsychInfo
<i>Follow-ups</i>	11	5	6	0
<i>Longitudinal</i>	7	5	2	0
<i>Quasi-experiments</i>	10	9	0	1
<i>Experiments</i>	3	3	0	0
<i>Multiple regression</i>	7	4	3	0
<i>Couldn't tell</i>	1	0	0	1

Table 4

Table Shell for Summarizing Year-1 Evaluation Findings
(Hypothetical Data)

Study	Standardized Effect Sizes	
	General Standardized Test (e.g., SAT-10 Mathematics)	Content Specific Test (e.g., ETS Algebra Test)
General Purpose TTPD	.15	-.05
Content-focused TTPD	.20	.40
Data-Driven Decision-making Tool PD	.25	.18

Figure 1. Logic Model for Technology-related Teacher Professional Development (TTPD)

