

[Not to be quoted without the expressed written consent of the author.]

**Using Sustained Professional Development as a Catalyst for
Improving Middle School Mathematics Performance:**

**Evidence from
Maine's Impact Study of Technology in Mathematics (MISTM)**

**A U.S. Department of Education Technology
Challenge grant (2003-2006)
PR#5318A03005**

**Dr. David L. Silvernail
Director
Maine Education Policy Research Institute
University of Southern Maine
(davids@usm.maine.edu)**

**March 2008
Paper Presented at the
Annual Meeting of the American Educational Research Association**

**Using Sustained Professional Development as a Catalyst for Improving Middle School
Mathematics Performance: Evidence from
Maine's Impact Study of Technology in Mathematics (MISTM)**

David L. Silvernail

**Maine Education Policy
Research Institute**

Introduction

The purpose of Maine's Impact Study of Technology in Mathematics (MISTM) was to investigate the impact of a sustained technology-infused teacher professional development program on student mathematics achievement. MISTM was designed to address a middle school students' mathematics achievement need in Maine. Analysis of results from the 2002 statewide Maine Education Assessment (MEA) indicated that Maine grade 8 students were failing to achieve the high academic expectations in mathematics outlined in Maine's learning standards. Statewide, 79% of all Maine 8th grade students failed to meet the Maine mathematics standard, and only 1% achieved the highest proficiency level. In addition, the 2002 MEA scores in mathematics showed that of the lowest performing Maine middle schools, 85% were rural and, in some of these schools, none of the grade 8 students met proficiency levels on the mathematics standards.

The specific technology infused sustained professional development program that was explored in this study was a program designed to take advantage of Maine's middle school laptop program. Beginning with the 2002-03 school year, and in all subsequent years, Maine has provided all grade 7 and 8 students and their teachers with laptop computers. All Maine's middle schools are wireless, permitting teachers and students to use their laptops throughout the school day and in a variety of settings and contexts. In addition, students may take their laptops home for use in the evenings, weekends, and on school vacations.

The program also has provided extensive technical assistance and targeted professional development programs to support the integration of the laptop program in all of Maine's 241 middle schools. School districts provide technical support to teachers and students, while a designated Teacher Leader at each middle school site assists teachers in integrating the laptops into their curriculum and instruction. Many types of professional development opportunities are provided to teachers as well. These include initial training, regional teacher leader meetings and content specialists meetings, interactive websites, and other statewide and local professional development activities relevant to technology. Finally, the Maine Department of Education

provides staff development personnel to assist schools and teachers in implementation of the laptop

The evaluation of the laptop program has provided evidence that, indeed, the introduction of the laptops in Maine's middle schools has impacted teaching and learning in many ways. However, it is widely recognized that access to technology tools alone will not ensure that all teachers in all disciplines will know how to use the technology to improve students' abilities to meet curriculum standards. This was evidenced in the Maine laptop program by evaluation data from the first year program which showed that only about 50% of the mathematics teachers reported using the laptops in their instruction, in contrast to 85% of teachers of other subject areas.

In light of the statewide MEA mathematics results and the evaluation findings on laptop use, Maine initiated several strategies to improve middle school mathematics teaching and learning. One strategy was to conduct systematic research on the impact of sustained professional development focused on mathematics. To that end, in 2003 Maine was awarded an Evaluating State Education Technology Program, (ESETP) grant from the United States Department of Education to conduct a randomized controlled trial (RCT) research study of the impact of a two-year technology-infused mathematics professional development program on student achievement. Partners with the Maine Department of Education in conducting the study were the Education Development Center (EDC) and the Maine Education Policy Research Institute (MEPRI).

What Works in Using Technology in Mathematics Classrooms

The potential of computers to enhance students' learning of mathematics has long been heralded, and a number of researchers have found positive effects of technology in teaching mathematics. For example, Heid & Blume (2003) provide a review of the research on technology to support the learning of algebra, organized in terms of instructional uses of technology tools for problem solving, communications, exploring and generalizing, representing, abstracting, and symbolizing. They found that technology may positively influence the development of algebraic concepts and lead to improved visual reasoning skills, and that technology enables the goals and content of algebra to shift from a procedural to a conceptual focus. These general conclusions are consistent with a review of the research on graphing calculators in secondary mathematics by Burrill et al (2002). They reported positive results in

that “students who use handheld graphing technology have a better understanding of functions, of variables, of solving algebra problems in applied contexts, and of interpreting graphs than those who did not use the technology” (p. v).

Further, in an analysis of different uses of technology in mathematics education based upon the 1996 NAEP data, Wenglinsky (1998) reported that “in essence, the study found that technology could matter, but that this depended upon how it was used” (p. 3). Specifically, for grade 8 Wenglinsky found that teacher professional development, the use of computers to teacher higher-order thinking skills, and the frequency of student home computer use were all positively related to achievement. He also found a serious equity issue: Black, poor, urban and rural students were less likely to be exposed to higher-order uses of computers and more likely to be exposed to lower-order uses than other students.

While the analyses by Heid & Blume (2003), Burrill et al (2002), Wenglinsky (1998) and others confirm that there is significant potential in the use of technology to enhance students’ learning of mathematics, they also identified many factors that have limited the actual impact of the technology on student learning, including: insufficient access to reliable computers, lack of quality software and of alignment of the software to the curriculum, teachers’ limited expertise in using technology to enhance teaching and learning, classroom management issues, and insufficient administrative support.

However, the prior research does not establish whether technology, with all the essential conditions for its effective integration into the curriculum, leads to increases in mathematical achievement by low performing middle school students in low-income rural communities in which there are an insufficient number of mathematics-credentialed teachers available. Conducting research based in low-performing rural schools within the Maine laptop project offered a unique opportunity to address this question and thereby provide a solid research base for future decisions about the educational benefits of technology in this type of context.

Principles of the Professional Development Approach

The intervention used in this experimental study was a technology-infused two-year professional development program. Researchers have established a number of key principles for effective professional development programs for K-12 educators. In a summary of these principles, Sparks and Hirsh (1997) described a “paradigm shift” in staff development, away from one-day in-service presentations to professional development as an integral, ongoing part of

teachers' lives, focused on improving student learning outcomes, based on inquiry into teaching and learning, and built upon interactions within professional learning communities. Major research studies and syntheses by Darling-Hammond & McLaughlin (1995); Ball & Cohen (1999); *National Staff Development Council*, (2001); *National Foundation for the Improvement of Education*, (1996) and others consistently agree that professional development is most effective when it:

- focuses on the curriculum standards for students and the alignment of teaching, curriculum, and assessment with those standards;
- fosters a deepening of subject-matter knowledge, a greater understanding of learning, and a greater appreciation of students' needs;
- centers around the critical activities of teaching and learning – planning lessons, evaluating student work, developing curriculum, improving classroom practices, and increasing student learning – rather than in abstractions and generalities;
- builds on investigations of practice through cases that involve specific problems of practice, questions, analysis, reflection, and substantial professional discourse;
- values and cultivates a culture of collegiality, involving knowledge, and experience-sharing among educators;
- is sustained, intensive, and continuously woven into the everyday fabric of the teaching profession, through modeling, coaching, and collaborations; and,
- begins with the mathematics content goals and effective teaching methods, rather than with the technology (when technology is employed).

The principles in the lists above informed the design of the professional development intervention that was implemented in this research study.

Study Design

Research Question and Logic Model

Based on the defined need to improve mathematics performance of Maine's middle school students, and the existing knowledge and research on effective professional development, and professional development using technology, this study was designed to answer a core research question:

Does a technology-infused sustained professional development program for teachers in rural schools result in significant improvements in middle school students' mathematics achievement?

A review of the extant literature was used to develop a logic model to guide the research. This Logic Model appears in Figure 1. As noted by Schmitz and Parsons (2005), "A logic

model’s purpose is to communicate the underlying ‘theory’ or set of assumptions or hypotheses...” of the program or research.

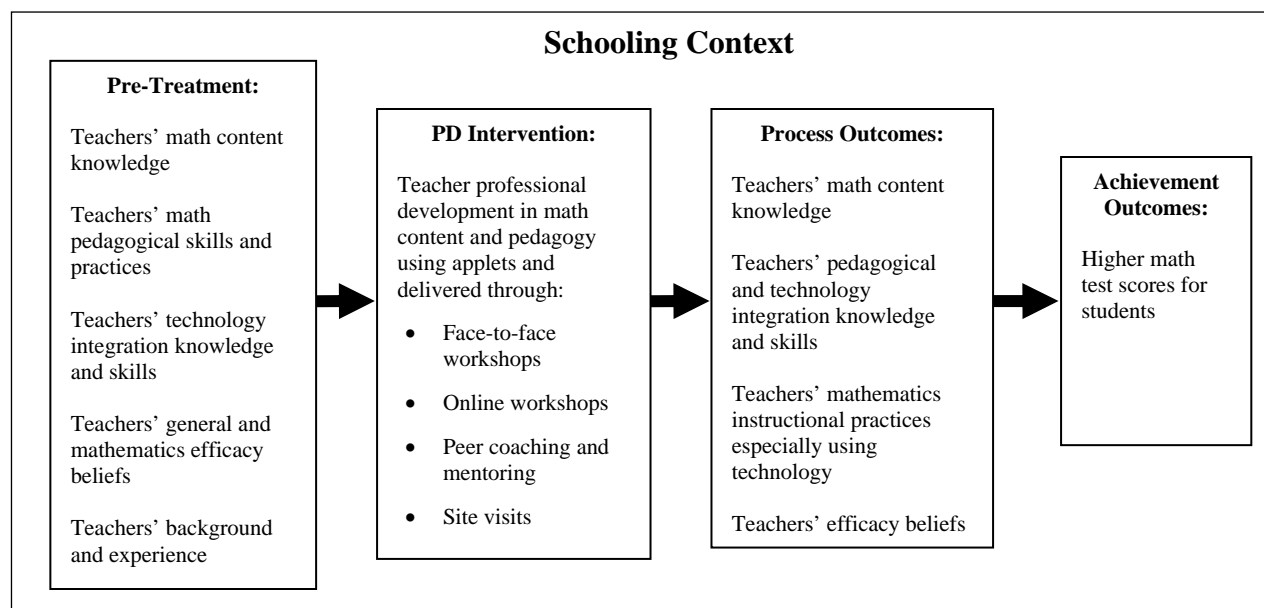


Figure 1: Logic Model for MISTM Research

The fundamental premise of this study was that teachers’ mathematics content knowledge and instructional practice may be directly linked to student performance. Changes are needed in both teachers’ content knowledge and pedagogical practices to improve students’ mathematical knowledge and understanding. Thus the logic underpinning this study was that a robust professional development intervention would result in changes in teachers’ content and pedagogical knowledge and skills, classroom practices, and beliefs about teaching. These changes would have a positive impact on students’ mathematics achievement.

A randomized control trial (RCT) research design was used in this study to explore the research question. The U.S. Department of Education is on record for supporting more so-called “scientifically based research,” and in recent years has encouraged giving funding priority to programs based on research which uses “an experimental design under which participants – e.g., students, teachers, classrooms, or schools – are randomly assigned to participate in the project activities being evaluated or to a control group that does not participate in the project activities being evaluated” (U.S. Department of Education, 2003a, p. 62446). More specifically a randomized pretest – post control group design was utilized in this study to determine the impact

of a two-year technology – infused professional development intervention program on middle school student mathematics performance. Figure 2 represents the research design.

Sample	Random Assignment of Schools & Treatment	Pretest Year One	Treatment	Posttest 1 Year One	Treatment Continued	Posttest 2 Year Two
Volunteer sample of purposive population of schools	Experimental Group	7 th Grade Teachers and Students	Professional Development Phase 1	7 th Grade Students	Professional Development Phase II	8 th Grade Teachers and Students
	Control Group	7 th Grade Teachers and Students	None	7 th Grade Students	None	8 th Grade Teachers and Students
Additional Data Collection Strategies: Teacher, Student, and Principal Surveys, Logs, Observations, and Case Studies						

Figure 2: Overview of Research Methodology, Randomized Pretest-Posttest control Group Design

Sample

The target population for generalization of the study findings was low performing rural Maine middle schools. Accordingly, the population of schools eligible for inclusion in the study was all Maine middle schools that:

1. served rural communities, as defined by the U.S. Census Bureau;
2. contained both 7th and 8th graders in the same building;
3. scored below the state average in mathematics MEA scores for the most recent two years; and,
4. had a significant proportion (at least 40%) of students eligible for free or reduced lunch programs.

The selection of the sample of schools for this study describe above involved several steps. All schools meeting the school population requirements were invited by the Maine Commissioner of Education to participate in the study. To be eligible for selection, all grade 7 and 8 teachers who taught mathematics at the school had to agree to complete the two-year professional development program if they were selected for the experimental treatment. This was an important stipulation in the study because it was hypothesized that students who were taught for two consecutive years by teachers participating in a sustained two-year professional development intervention program would significantly improve their mathematics performance. In addition, the participating schools had to agree to participate in all required data collection activities.

One hundred ninety-one (191) Maine middle schools were eligible for participation in the study. A total of 56 schools accepted the Commissioner of Education's invitation to participate in the study. After the principals, and all 7th and 8th grade teachers in a middle school, signed and returned a written agreement to the study conditions, the schools were randomly assigned to either the experimental or control groups. Random assignment was done at the school level to decrease the likelihood of sharing of resources, instructional strategies, etc., between the experimental group teachers and the control group teachers, and to decrease the occurrence of the John Henry Effect.

Twenty-eight (28) schools were randomly assigned to each study group (Experimental and Control). However, once the schools were assigned to the two groups, nine schools (7 Experimental schools and 2 Control schools) notified the researchers of their decision *not* to continue in the study, even though they had agreed to the participation stipulations. An analysis of school, teacher, and student characteristics of the schools that opted out from the original sample indicated that these schools were not significantly different from the remaining schools. Further analysis indicated there were no statistically significant differences between the remaining experimental and control schools based on school size ($t=1.643$; $df = 46$; $p>.05$) or teacher characteristics (education level, length of teaching experience, gender). MEA performance of the 8th graders in each school for the most recent two years revealed no significant difference in school wide MEA performance ($t = 1.918$; $df = 46$; $p>.05$), and there was no significant difference in eligibility levels for free or reduced lunch programs ($t=0.324$; $df = 46$; $p>.05$).

Treatment

The goals of the experimental intervention in this study were fourfold:

- Content – deepen teachers' mathematical content knowledge in the areas of *Numbers and Operations* and *Patterns* in Maine's statewide learning standards.
- Pedagogy – improve teachers' pedagogical practice in technology infused mathematics classrooms.
- Technology Integration – develop and apply strategies that support the integration of technology for the teaching, learning, and assessment of mathematics.
- Professional Learning Community – engage teachers in meaningful interaction and dialogue about mathematics through face-to-face and online environments.

The experimental intervention consisted of four interrelated professional development components. These were: (1) face-to-face workshops; (2) online workshops; (3) peer coaching and mentoring; and (4) site visits.

(1) *Face-to-face workshops*

The face-to-face workshops focused on (1) building teachers' knowledge of the key curriculum standards relevant to preparing students for proficiency in the Maine learning standards for mathematics; (2) deepening teachers' own understandings of the central mathematical ideas; (3) introducing the technology tools that are designed to help teach these ideas; (4) integrating work on formative assessments of students learning with the use of the technology tools to support the identified learning needs; (5) working with the teachers on plans to integrate the tools into the curriculum to help students meet the identified mathematics standards; and (6) building a shared vision with building administrators regarding technology integrated mathematics instruction.

(2) *Online Workshops*

Following the first set of face-to-face workshops, an online workshop was provided, comprised of ten weeks of activities. Teachers were expected to spend five hours per week on workshop activities, both online and offline. Teachers participated in the online and face-to-face workshops as members of the same group, with the same workshop leaders. The online workshops continued work begun in the first face-to-face workshops, extending it to additional content standards, mathematical concepts and skills, and software tools.

(3) *Peer Coaching and Mentoring*

The third component of the professional development approach was the establishment of a system of peer coaching among the participating teachers and regional mathematics content leaders. Building upon the approach recommended by Loucks-Horsley, Hewson, Love & Stiles (1998), the coaching model provided opportunities for the participating educators to build upon their own strengths. The model was a collegial peer-coaching model where teachers were paired up with a peer who was assigned to teach the same grade level at a nearby school. They supported one another, problem-solved, shared strategies, observed and gave feedback to one another.

(3) Site Visits

The fourth component consisted of site visits. The visits were designed to support the training and mentoring with on-going observation and feedback as teachers implemented new learning, resources and strategies into the classroom. This component was designed to provide continuous improvement and capacity building over the course of the two year intervention.

Overall, teachers received 30 hours of face-to-face and 50 hours of online professional development each year of the two year intervention. Additionally, participating teachers received at least 24 additional hours each year of ongoing support via peer coaching, site visits, and the online forum and resources, to prepare them to improve their instructional practices and use technology effectively. The total professional development contact was 208 hours over the course of two years.

Instrumentation

A series of data collection tools were used to document and analyze the impacts of the professional development intervention. These tools were specifically designed for this research study.

Process Measures

A series of student and teacher surveys and teacher logs were used throughout the study to collect process measures. The data collection instruments were designed by the research team, and administered in an online environment whenever possible.

Teacher Surveys: Four (4) teacher surveys were administered over the two year intervention. The surveys collected a variety of information, including: (1) teacher demographics; (2) technology use levels; (3) mathematics instructional strategies; (4) teacher efficacy; (5) teaching philosophy.

Teacher Logs: Teachers completed four electronic logs over the course of the two year intervention, and submitted these electronically to the research team. Two logs focused on the use of various instructional strategies used by the teachers. The remaining two logs focused on the use of technology in teaching mathematics.

Student Surveys: Three (3) student surveys were administered in the two years. The focus of the surveys was to collect students' perceptions of how they were being asked by their teachers to use laptops in the classroom and outside of the classroom. Categories of survey items included:

(1) student demographics; (2) perceptions of technology competency; (3) technology beliefs; (4) uses of technology.

Student Achievement

The mathematical focus of the research study was developing students' readiness to succeed in algebra. The Maine learning standards, called Learning Results, are structured around four major mathematics clusters: (1) Numbers and Operations; (2) Shape and Size; (3) Mathematical Decision-Making; and (4) Patterns. Each cluster contains a series of performance indicators designed to guide instruction and assessment. Two clusters (1. Numbers and Operations & 4. Patterns) were the focus of this research study. An analysis of MEA mathematics performance over four years indicated that student performance in these two performance indicator clusters was lowest on the statewide tests.

Three performance indicators in these two clusters were selected for study:

Cluster 1: Numbers and Operations

Performance Indicator: **A1.** Use numbers in a variety of equivalent and interchangeable forms (e.g., integer, fraction, decimal, percent, exponential and scientific notation) in problem solving.

Cluster 4: Patterns

Performance Indicator: **G1.** Describe and represent relationships with tables, graphs, and equations.

K2. Use statistics, tables, and graphs to communicate ideas and information in convincing presentations and analyze presentations of others for bias or deceptive presentation.

The research team conducted an analysis of the existing statewide 8th grade MEA mathematics test to determine its viability for assessing student achievement in the mathematics focus areas of this study. It was determined that the statewide test was not viable for two primary reasons. First, there were an insufficient number of items on the test to obtain valid and reliable test scores in the three performance areas. Second, in any given year there was no guarantee that the three specific performance indicators which were the targets in this study would be assessed.

Given this analysis of the statewide MEA mathematics test, the research team then communicated with other investigators across the United States and collected and examined existing instruments that might be used for assessing students' as well as teachers' knowledge on these specific performance indicators. The team concluded that no existing instruments would

meet the project’s needs. The team also considered how best to administer each of the assessments, using either paper-and-pencil or online formats. Because every middle school student and teacher in Maine has a laptop computer, and because the MEA was scheduled to be offered online in the future, the team decided that the G1/K2 portion of the student assessment and the entire teacher assessment would be delivered online.

Two separate instruments were developed for students. One focused on performance indicator A1. The other focused on performance indicators G1 and K2 together. Each of the assessments was designed to take no more than 45 minutes, and the online assessment (G1/K2) was designed to be accessible through an internet browser. Table 1 describes the student assessment instruments.

Table 1: Student Assessment Instruments

Cluster	Standard Label	Performance Indicators (Middle Grades 5-8)	Type of Test Item
Maine Learning Results- Content Cluster 1: <i>Numbers & Operations</i>	A. Numbers & Number Sense	A1: Use numbers in a variety of equivalent and interchangeable forms in problem solving. (eg., integer, fraction, decimal, percent, exponential, & scientific notation)	5 Multiple Choice 7 Short Answer 2 Constructed Response
Maine Learning Results- Content Cluster 4: <i>Patterns</i>	G. Patterns, Relations, Functions K. Mathematical Communication	G1: Describe and represent relationships with tables, graphs, & equations. K2: Use statistics, tables & graphs to communicate ideas & information in convincing presentations & analyze presentations of others for bias or deceptions.	5 Multiple Choice 11 Short Answer 3 Constructed Response

The student assessments were designed by researchers at the Education Development Center with assistance from a mathematics specialist at the Maine Department of Education. Over the course of the study, three different versions of each test were developed, field-tested, validity and reliability characteristics. Standard procedures for establishing the validity and reliability of the assessment instruments were used in examining test items, including principle component and reliability analyses, and procedures to examine item difficulty, and item discrimination, and to check for item biases. These analyses resulted in establishing that the student assessment forms were valid and reliable for research purposes.

Teacher Achievement

The teacher assessments were also designed to assess teachers’ content and pedagogical knowledge in A1, and G1 and K2. Pedagogical knowledge included understanding students’ mathematical thinking, as well as understanding how to effectively build upon and develop mathematical thinking. The teacher assessments were designed to assess teachers’ knowledge in

a new way. A review of available teacher tests indicated no existing tests which measured teachers' knowledge in the specific content areas which were the focus of this research in sufficient detail, nor were existing tests in a format the research team believed acceptable to administer to experienced teachers. Accordingly, a new format was used in developing the teacher assessments. The teacher assessments included a number of incorrect responses from students to pilot items, and teachers were asked to explain the students' thinking and suggest questions that would help the students to better understand the mathematics.

Data Analysis Procedures

To examine the effects of the technology-infused professional development on student achievement, primary and secondary analysis were conducted. Analysis of variance statistics were employed in examining the core research question. Independent t-tests were used to examine the equivalence of the experimental and control groups, and dependent t-tests were used in examining pretest to posttest gains in the project designed mathematics tests. ANOVA and ANCOVA procedures were used in examining main effects and interactions on the grade 8 MEA, and effect sizes were also calculated.

To examine further the effect of the two-year sustained technology-infused professional development and instructional use of technology tools on student achievement, hierarchical linear modeling (HLM) techniques were used to model differences in achievement between the experimental and control groups. By using a two-level model which takes into account the nested data structure, the relationship between student achievement and membership in the experimental and control groups could be better estimated. Student knowledge, including measures of previous achievement as well as teacher knowledge and skill characteristics and context variables were incorporated into the two-level models to more closely examine the effects of the technology-rich professional development on student achievement. In the two-level model, students within a teacher's classroom were assigned to level-1, and teachers were assigned to level-2.

A final tool used to understand the possible relationships between variables was path analysis. More specifically, path analysis was used to test the fidelity of the logic model in representing the relationships between group membership, the intervention, and the impacts of the intervention on teachers' knowledge, beliefs and practices, and student achievement.

Results

Table 2 reports the instrument data collection schedule and the percentages of usable return rates over the course of the two-year study. As may be seen from the table, return rates at individual data collection points were very high over the two-year intervention period. However, it is important to note that some teachers and students left the study for various reasons (e.g., moved to another school district, teacher retirement, teacher leaving profession, etc.), while others joined the study after the intervention had begun. Consequently, although the return rates were very high they do not represent the same teachers and students in all cases.

Table 2: Schedule of Data Collection

Instrument	Year 1 Fall	Year 1 Spring	Year 2 Fall	Year 2 Spring
Teacher Assessment	100%	-	-	94%
Teacher Surveys	98%	93%	-	95%
Teacher Logs	99%	95%	95%	93%
Student Assessments	93%	91%		93%
Student Surveys	82%		76%	73%

As described earlier, the original sample of 56 volunteer schools were randomly assigned to experimental and control groups. And even though schools had agreed to participate in the study regardless of their group assignment, seven schools (7 Experimental schools and 2 Control school) chose not to participate in the study before the study began implementation. Thus, the starting sample of schools was 21 Experimental schools and 27 Control schools with 57 experimental group teachers and 54 control group teachers. Analysis of teacher demographics revealed that the two groups of teachers were equivalent on demographic characteristics such as education level, overall teaching experience, and years of experience teaching mathematics.

The sample was further reduced over the two-year study period. Table 3 presents the overall samples at the beginning and end of the intervention. As may be seen from the table, the teacher sample at the end of the intervention period was 91 teachers, a number equal to 83% of the beginning teacher sample. In the case of the student sample the ending sample was equal to 87% (n=4671) of the beginning sample.

Table 3: Overall Beginning and Ending Samples

Time	Schools		Teachers		Students	
	Exp.	Control	Experimental	Control	Experimental	Control
Beginning	21	27	55	57	2507	2860
Spring 2006 Usable Sample	19	27	40	51	1946	2725

However, the sample used in addressing the research question had to be further reduced because of lost data points, and implementation fidelity issues. That is to say, because of the realities of schools, complete data was not obtained for all teachers and students. For example, some teachers left employment in the study schools, some teachers did not complete all study instruments, and students moved into and out of the study schools, and /or did not complete all the assessments. For purposes of answering the research question it was important to focus the analysis on those teachers deemed to have participated at a high level in the intervention. That is, teachers who had participated continuing in the two-year professional development activities, and teachers who had implemented the intervention strategies at a high level of fidelity in their classrooms. Thus the study sample was further reduced and these samples appear for Time 3 and 4 in Table 4.

Time 3 represents the number of teachers who participated continuous for 20 or more months of the two-year intervention, and teachers who had submitted sufficient data to be included in the study sample (n=44). The Time 3 sample was further reduced for purposes of addressing the study question.

Table 4: Reduction in Sample Size

Time	Schools		Teachers		Students	
	Exp.	Control	Exp.	Control	Exp.	Control
3. 20+ months Continuous Involvement	14	20	21	43	521	775
4. Final Study Sample (FSS)	7	17	10	27	281	692

All the experimental group teachers implemented the intervention activities in light of various factors such as their own teaching efficacy, school and classroom conditions, and the mathematics curriculum used in their schools. Assessments by the intervention providers were

used in classifying the 21 experimental group teachers on a scale of 0 (low) to 4 (high) as implementers of intervention strategies (e.g., completion of intervention activities, use of intervention materials in their classes, etc.). Teachers with assessments of 3 or 4 were selected for use in the final analysis of results. Thus, the 10 teachers and their students were labeled as the Final Study Sample (FSS); the group used in the strictest analysis of the core research questions. That is to say, the study was designed to learn if two-years of sustained professional development with teachers had a positive impact on student learning. Accordingly, the FSS became those teachers who; (1) had participated in the intervention professional development program for 20 – 24 months; (2) had implemented intervention activities in their classrooms; and, (3) had provided pre- and post-intervention data on themselves and their students. Put another way, the FSS included only students who were in classrooms, both in the 7th and 8th grades, and who were taught in these two grades by teachers who participated in the two-year professional development program. The resulting sample in this FSS became 37 teachers (10 experimental teachers and 27 control teachers). The FSS experimental sample became smaller than the control sample because the experimental teachers had to have completed the intervention activities and received an implementation assessment of 3 or 4 (Meets or Exceeds criteria).

In order to determine the comparability of the Beginning Sample and the Final Study Sample (FSS), three basic teacher demographic profiles were examined. Both samples were very similar in total years of teaching experience, with the FSS Group teachers with a few more years experience teaching mathematics and slightly higher education degree levels. However, statistical analysis of the profiles revealed no statistically significant differences.

For the initial core hypothesis testing total student group performance was examined at two points in time; at the end of one year of intervention and after two years of intervention. Given the random assignment used in this study, the importance of nested variables was recognized and explored following the initial hypothesis testing. The initial testing was conducted merely to explore overall effects of the intervention strategy. The core hypothesis was as follows:

H₀: There will be no significant differences in middle school student mathematics achievement for experimental and control group classroom students.

Table 5 reports the student achievement scores at the beginning and end of the two year intervention for the Final Study Sample (FSS) only. As the results indicate, for the FSS, experimental group classroom students and control group classroom students did differ in Fall

Table 5: Student Total Test Score Results After Two Year Intervention

Content		Experimental (n=281)	Control (n=692)	t=	p=	Effect Size
Mathematics Test Total Score	Fall 2004	32.1%	27.8%	3.80	<.01	0.29
	Spring 2005	54.6%	47.9%	3.62	<.01	0.39

2004. But an analysis of covariance (ANCOVA) for group effects indicated overall test score results were also significantly different at the end of the two year intervention, in favor of the experimental group students. Overall, the experimental group students gained more over the two years in which their teachers participated in the sustained technology-infused professional development program.

To further examine these achievement results, test scores were analyzed by subtests (A1 and G1/K2) and these results appear in Table 6. This additional analysis uncovered some differences in performance. For both subtests, there were statistical differences in the pretest (Fall 2004) and analysis of covariance (ANCOVA) revealed A1 posttest score results (Spring 2006) for the experimental group students did not differ significantly from the scores of control group students. However, in the case of the G1/K2 test score results, the difference between groups were statistically significant. Experimental group students scored significantly higher.

Table 6: Student Subtest Score Results After Two Year Intervention

Subtest Score Results						
Content		Experimental (n=281)	Control (n=692)	t=	p=	Effect Size
A1 Subtest	Fall 2004	30.1%	25.8%	3.87	<.01	0.28
	Spring 2006	56.0%	51.5%	0.35	>.01	0.22
G1/K2 Subtest	Fall 2004	35.4%	31.2%	3.30	<.01	0.24
	Spring 2006	53.4%	44.8%	5.97	<.01	0.50

Thus, while the total score analysis would lead one to reject the null hypothesis, the subtest score analysis suggested that the null hypothesis should be accepted for A1 results and rejected for G1/K2 results.

The preliminary analysis of the results suggested overall group effects in favor of the experimental group students for the G1/K2 content, but not in the case of the A1 content. These group level results were examined further in a second phase of the analysis using hierarchical linear modeling (HLM) techniques. HLM provided the opportunity to examine the nested nature of the data; teachers nested in schools randomly assigned to experimental and control groups, and students nested within classrooms.

Hierarchical linear modeling was used to address the core research question regarding the effects of the sustained teacher professional development on student outcomes. In addition to its main purpose, it provided two additional types of information. First, variance was partitioned into within- and between-classroom components, and a determination of the percentage of each of these components is explained by the predictors. Second, the strength of the conditional associations between other teacher and student variables and student outcomes was assessed.

The organizational structure of the sample is rather complex in that students are assigned to classes which are taught by teachers who work within schools. Further, students are assigned a mathematics teacher in each of the two grades. Groups were assigned to the treatment or control group at the school level. Ideally, even ignoring multiple classes taught by the same teacher, one would have a three level model, students within teachers within schools, cross-classified at the teacher level. However, the quantity of data was not sufficient to support stable estimates of three-level models with predictors. Preliminary analysis suggested much more variation between teachers than between schools, perhaps due in part to a number of schools having only one mathematics teacher in each grade. In fact, in a three-level model of students within teachers within students, the variance at the school level was not statistically significant. Therefore, two-level models of students within 8th grade teachers were created. Student-level predictors are centered around the group mean. Teacher and school variables are centered around the grand mean. Variables examined in the HLM included selected teacher characteristics and one student characteristic (poverty).

Key teacher variables included in the HLM analysis included teacher content and pedagogical knowledge, teaching philosophy, and technology use. For each of the student outcome variables, total mathematics score (A1 & G1/K2), A1 score, and G1/K2 score, a hierarchical linear model with no predictors was estimated followed by four models with predictors. The model with no predictors was used to break out the total variance into between-

teacher and within-teacher components. Pre-test and other prior characteristics were added successively in the first three models. Student pre-test was entered first, followed by teacher and school characteristics. Then treatment group assignment is entered in order to test for group effects on student outcomes directly. In a fourth model, to explore the relation between current teacher qualities and student outcomes, the predictors included post-professional development teacher knowledge and characteristics.

The results of the four hierarchical linear models of total mathematics score are provided in Table 7. According to the model with no predictors, as seen at the bottom of the table, a preponderance of the overall variation in student scores, 71.6%, was within teacher classrooms compared to 28.4% between teacher classrooms. In Model 1, individual student pretest scores were predictors of individual posttest scores and teacher classroom mean pretest scores were predictors of teacher classroom mean posttest. The estimated coefficients for both predictors were statistically significantly different from zero. Most of the variation between teacher classrooms, 83.1%, was explained by this model as well as much of the variation within teacher classrooms, 47.0%. Of the unexplained variance, only 11.3% (i.e., $0.00164 \div 0.01457$) was between teacher classrooms.

Additional teacher and school variables were entered as predictors of teacher classroom mean scores in Model 2. These were pre-tests and other variables representing the state of affairs before implementation of the two year sustained professional development intervention. However, none of the additional variables had statistically significant coefficients. Model 3 included group membership as a predictor. Neither participation in the strictest group, the FSS group, which represented full participation in the treatment group for the length of the study, nor participation in the treatment group without meeting the criteria for being identified in the strictest group was a significant predictor of student scores.

Whereas Model 3 treated student scores as a function of characteristics of students and teachers before professional development along with indicators representing participation and degree of participation in the treatment group, Model 4 treated student scores as a function of student pre-tests and characteristics of teachers after participation in the professional development intervention. Student pretest scores and Teacher posttest scores were statistically significant predictors of student posttest scores.

The student assessment scores were analyzed further by examining subtest scores. Four models for student A1 assessment scores were estimated. The results are shown in Table 8. Only student pretest scores were identified as statistically significant predictors. Individual student pretest scores were predictive of posttest scores and classroom mean pretest scores were predictive of classroom mean posttest scores. No teacher variable, including treatment group membership and teacher knowledge pre- or post-training, was shown to be related to student outcomes. No additional variance was explained by adding teacher and school level predictors.

Table 7: Total Mathematics Test Scores Two-Year Two-Level Analysis

Outcome = Student Score (A1 & G1)		Model 1		Model 2		Model 3		Model 4			
		Coeff.	Sig.	Coeff.	Sig.	Coeff.	Sig.	Coeff.	Sig.		
Student Level Predictor (n=1,456)								Student Level Predictor			
Fall 04 Student Assessment		0.816018	0.000	0.816018	0.000	0.816018	0.000	Fall 04 Student Assessment	0.816018	0.000	
Teacher and School Level Predictors (n=51)								Teacher and School Level Predictors			
<i>(intercept)</i>		<i>0.491784</i>	<i>0.000</i>	<i>0.490815</i>	<i>0.000</i>	<i>0.490973</i>	<i>0.000</i>	<i>(intercept)</i>	<i>0.491037</i>	<i>0.000</i>	
Mean Fall 04 Student Assessment		0.932184	0.000	0.856281	0.000	0.853572	0.000	Mean Fall 04 Student Assessment	0.859571	0.000	
Fall 04 Teacher Assessment				0.150415	0.109	0.159063	0.105	Spring 06 Teacher Assessment	0.137999	0.047	
NSLP (School)				-0.000684	0.125	-0.000683	0.138	NSLP (School)	-0.000504	0.254	
Teaching Philosophy (Pre)				0.004984	0.564	0.004225	0.645	Teaching Philosophy (Post)	0.011709	0.125	
Laptop Use (Pre)				0.006900	0.879	0.001110	0.814	Laptop Use (Post)	-0.003734	0.279	
Conceptual Use (Pre)				-0.008915	0.362	-0.009449	0.361	Use of Applets (Post)	-0.001221	0.874	
Strictest Treatment Group						0.003849	0.846	Strictest Treatment Group			
Treatment Group, Not Strictest						0.007712	0.679				
Variance Component	No Predictors		Model 1		Model 2		Model 3		Variance Component	Model 4	
	Total Variance	Percent of Total	Residual Variance	Variance Explained	Residual Variance	Variance Explained	Residual Variance	Variance Explained		Residual Variance	Variance Explained
Between Teacher	0.00969	28.4%	0.00164	83.1%	0.00155	84.0%	0.00163	83.2%	Between Teacher	0.00139	85.7%
Within Teacher	0.02440	71.6%	0.01293	47.0%	0.01291	47.1%	0.01292	47.0%	Within Teacher	0.01292	47.0%
Total	0.03409	100.0%	0.01457	57.3%	0.01446	57.6%	0.01455	57.3%	Total	0.01431	58.0%

Table 8: Total Mathematics A1 Test Scores Two-Level Analysis

Outcome = Student Score (A1)		Model 1		Model 2		Model 3		Model 4			
		Coeff.	Sig.	Coeff.	Sig.	Coeff.	Sig.	Coeff.	Sig.		
Student Level Predictor (n=1,456)								Student Level Predictor			
Fall 04 Student Assessment		0.909224	0.000	0.909224	0.000	0.909224	0.000	Fall 04 Student Assessment	0.909224	0.000	
Teacher and School Level Predictors (n=51)								Teacher and School Level Predictors			
<i>(intercept)</i>		<i>0.524399</i>	<i>0.000</i>	<i>0.524519</i>	<i>0.000</i>	<i>0.525015</i>	<i>0.000</i>	<i>(intercept)</i>	<i>0.524356</i>	<i>0.000</i>	
Mean Fall 04 Student Assessment		1.041025	0.000	1.010550	0.000	1.009470	0.000	Mean Fall 04 Student Assessment	1.028633	0.000	
Fall 04 Teacher Assessment				0.090416	0.277	0.111829	0.196	Spring 06 Teacher Assessment	0.018646	0.806	
NSLP (School)				-0.000083	0.862	-0.000020	0.968	NSLP (School)	-0.000019	0.970	
Teaching Philosophy (Pre)				-0.000607	0.945	-0.003303	0.714	Teaching Philosophy (Post)	0.003021	0.720	
Laptop Use (Pre)				0.002779	0.560	0.004147	0.400	Laptop Use (Post)	-0.000918	0.807	
Conceptual Use (Pre)				-0.013066	0.189	-0.015016	0.144	Use of Applets (Post)	0.004254	0.613	
Strictest Treatment Group						0.019482	0.331	Strictest Treatment Group			
Treatment Group, Not Strictest						0.019866	0.310	Treatment Group, Not Strictest			
Variance Component	No Predictors		Model 1		Model 2		Model 3		Variance Component	Model 4	
	Total Variance	Percent of Total	Residual Variance	Variance Explained	Residual Variance	Variance Explained	Residual Variance	Variance Explained		Residual Variance	Variance Explained
Between Teacher	0.01080	25.3%	0.00147	86.4%	0.00153	85.8%	0.00155	85.6%	Between Teacher	0.00166	84.6%
Within Teacher	0.03197	74.7%	0.01706	46.6%	0.01706	46.6%	0.01706	46.6%	Within Teacher	0.01707	46.6%
Total	0.04277	100.0%	0.01853	56.7%	0.01859	56.5%	0.01861	56.5%	Total	0.01873	56.2%

Four similar models were estimated for student G1/K2 assessment scores. The results are shown in Table 9. A greater number of statistically significant predictors were identified for G1/K2 than for A1 scores. Student pretest scores, teacher pretest scores, and the percentage of students eligible for free or reduced price lunches were all predictive of posttest scores, as seen in Model 2 and Model 3. In Model 4, student pretest, teacher posttest, and teaching philosophy were predictive of student posttest scores. The variance components were similar for A1, G1/K2 and Average A1 G1/K2 scores. Most of the variation in student posttest scores was within classrooms, and of the remaining variation, that between classrooms, most are explained by variations in student pretest scores.

Phase 3 of the analyses consisted of several path analyses to further explore the relationships between variables and to test the logic model underpinning the study. First, bivariate correlations were examined between key study variables, and these correlations appear in Table 10. Analysis of this simple correlation table suggested some of the relationships that might be found in the path analyses. Student poverty, as measured by eligibility for free or reduced lunch, was not generally related to pretest student scores, but were related to posttest student scores. Students' pretest and posttest scores were highly correlated, and students' posttest scores were moderately correlated with teacher pre and posttest content knowledge. Teaching philosophy was not related to student performance nor was post-intervention laptop use levels.

Table 9: Total Mathematics G1 Test Scores Two-Level Analysis

Outcome = Student Score (G1)		Model 1		Model 2		Model 3		Model 4			
		Coeff.	Sig.	Coeff.	Sig.	Coeff.	Sig.	Coeff.	Sig.		
Student Level Predictor (n=1,456)								Student Level Predictor			
Fall 04 Student Assessment		0.491731	0.000	0.491731	0.000	0.491731	0.000	Fall 04 Student Assessment	0.491731	0.000	
Teacher and School Level Predictors (n=51)								Teacher and School Level Predictors			
<i>(intercept)</i>		<i>0.459735</i>	<i>0.000</i>	<i>0.457627</i>	<i>0.000</i>	<i>0.457807</i>	<i>0.000</i>	<i>(intercept)</i>	<i>0.458332</i>	<i>0.000</i>	
Mean Fall 04 Student Assessment		0.721769	0.000	0.602377	0.000	0.596150	0.000	Mean Fall 04 Student Assessment	0.636313	0.000	
Fall 04 Teacher Assessment				0.229112	0.027	0.229575	0.031	Spring 06 Teacher Assessment	0.219917	0.002	
NSLP (School)				-0.001359	0.013	-0.001358	0.015	NSLP (School)	-0.000874	0.088	
Teaching Philosophy (Pre)				0.009187	0.386	0.008824	0.424	Teaching Philosophy (Post)	0.020841	0.023	
Laptop Use (Pre)				-0.001208	0.827	-0.000891	0.878	Laptop Use (Post)	-0.006375	0.116	
Conceptual Use (Pre)				-0.011708	0.343	-0.012577	0.338	Use of Applets (Post)	-0.002730	0.762	
Strictest Treatment Group						0.006218	0.806	Strictest Treatment Group			
Treatment Group, Not Strictest						0.003180	0.890	Treatment Group, Not Strictest			
Variance Component	No Predictors		Model 1		Model 2		Model 3		Variance Component	Model 4	
	Total Variance	Percent of Total	Residual Variance	Variance Explained	Residual Variance	Variance Explained	Residual Variance	Variance Explained		Residual Variance	Variance Explained
Between Teacher	0.00939	25.6%	0.00309	67.1%	0.00229	75.6%	0.00244	74.0%	Between Teacher	0.00177	81.2%
Within Teacher	0.02729	74.4%	0.02090	23.4%	0.02087	23.5%	0.02088	23.5%	Within Teacher	0.02089	23.5%
Total	0.03668	100.0%	0.02399	34.6%	0.02316	36.9%	0.02332	36.4%	Total	0.02266	38.2%

Table 10: Correlations of Key Variables

	Student Scores (Pre)			Teacher Scores (Pre-)			Teacher Other (Pre)		Poverty	Student Scores (Post)			Teacher Scores (Post)			Teacher Other (Post)	
	Mean Student Score F'04 A1 G1	Mean Student Score F'04 A1	Mean Student Score F'04 G1	Teacher Score F'04 A1 G1	Teacher Score F'04 A1	Teacher Score F'04 G1	Teaching Philosophy (Pre)	Laptop Use (Pre)	Free/Reduced Lunch Percentage	Mean Student Score S'06 A1 G1	Mean Student Score S'06 A1	Mean Student Score S'06 G1	Teacher Score S'06 A1 G1	Teacher Score S'06 A1	Teacher Score S'06 G1	Teaching Philosophy (Post)	Laptop Use (Post)
Mean Student Score F'04 A1 G1	1	.933(**)	.946(**)	0.237	0.11	.293(*)	-0.15	-0.118	-0.25	.888(**)	.885(**)	.835(**)	0.258	0.267	0.213	-0.015	0.067
Mean Student Score F'04 A1	.933(**)	1	.767(**)	0.231	0.111	.277(*)	-0.113	-0.068	-.338(*)	.897(**)	.916(**)	.819(**)	0.249	0.262	0.202	-0.092	0.086
Mean Student Score F'04 G1	.946(**)	.767(**)	1	0.216	0.096	0.275	-0.167	-0.149	-0.141	.778(**)	.755(**)	.753(**)	0.237	0.241	0.199	0.056	0.042
Teacher Score F'04 A1 G1	0.237	0.231	0.216	1	.880(**)	.894(**)	0.173	.453(**)	-0.219	.388(**)	.311(*)	.445(**)	.697(**)	.647(**)	.634(**)	0.251	.284(*)
Teacher Score F'04 A1	0.11	0.111	0.096	.880(**)	1	.579(**)	0.198	.393(**)	-0.264	0.273	0.225	.306(*)	.581(**)	.553(**)	.517(**)	0.238	0.258
Teacher Score F'04 G1	.293(*)	.277(*)	0.275	.894(**)	.579(**)	1	0.104	.424(**)	-0.123	.395(**)	.311(*)	.460(**)	.625(**)	.567(**)	.579(**)	0.207	0.243
Teaching Philosophy (Pre)	-0.15	-0.113	-0.167	0.173	0.198	0.104	1	.426(**)	0	-0.047	-0.065	-0.024	0.179	0.188	0.145	.500(**)	0.131
Laptop Use (Pre)	-0.118	-0.068	-0.149	.453(**)	.393(**)	.424(**)	.426(**)	1	-0.077	0.016	-0.061	0.096	.382(**)	.410(**)	.303(*)	.383(**)	.519(**)
Free/Reduced Lunch Percentage	-0.25	-.338(*)	-0.141	-0.219	-0.264	-0.123	0	-0.077	1	-.380(**)	-.360(**)	-.377(**)	-.290(*)	-0.237	-.290(*)	-0.033	-0.167
Mean Student Score S'06 A1 G1	.888(**)	.897(**)	.778(**)	.388(**)	0.273	.395(**)	-0.047	0.016	-.380(**)	1	.971(**)	.967(**)	.399(**)	.363(**)	.370(**)	-0.039	0.173
Mean Student Score S'06 A1	.885(**)	.916(**)	.755(**)	.311(*)	0.225	.311(*)	-0.065	-0.061	-.360(**)	.971(**)	1	.878(**)	.310(*)	.292(*)	.279(*)	-0.071	0.1
Mean Student Score S'06 G1	.835(**)	.819(**)	.753(**)	.445(**)	.306(*)	.460(**)	-0.024	0.096	-.377(**)	.967(**)	.878(**)	1	.468(**)	.415(**)	.442(**)	-0.003	0.24
Teacher Score S'06 A1 G1	0.258	0.249	0.237	.697(**)	.581(**)	.625(**)	0.179	.382(**)	-.290(*)	.399(**)	.310(*)	.468(**)	1	.898(**)	.935(**)	0.27	0.262
Teacher Score S'06 A1	0.267	0.262	0.241	.647(**)	.553(**)	.567(**)	0.188	.410(**)	-0.237	.363(**)	.292(*)	.415(**)	.898(**)	1	.683(**)	0.268	.361(**)
Teacher Score S'06 G1	0.213	0.202	0.199	.634(**)	.517(**)	.579(**)	0.145	.303(*)	-.290(*)	.370(**)	.279(*)	.442(**)	.935(**)	.683(**)	1	0.232	0.144
Teaching Philosophy (Post)	-0.015	-0.092	0.056	0.251	0.238	0.207	.500(**)	.383(**)	-0.033	-0.039	-0.071	-0.003	0.27	0.268	0.232	1	.293(*)
Laptop Use (Post)	0.067	0.086	0.042	.284(*)	0.258	0.243	0.131	.519(**)	-0.167	0.173	0.1	0.24	0.262	.361(**)	0.144	.293(*)	1

*Significant at .05 level; **Significant at .01 level

To further examine relationships, a series of six path analyses were created, and these appear in Figures 3-8 on the following pages. Figures 4 and 5 display the path analysis for total mathematics student performance. Darker lines and arrows signify statistically significant partial correlation at the minimum .05 level.

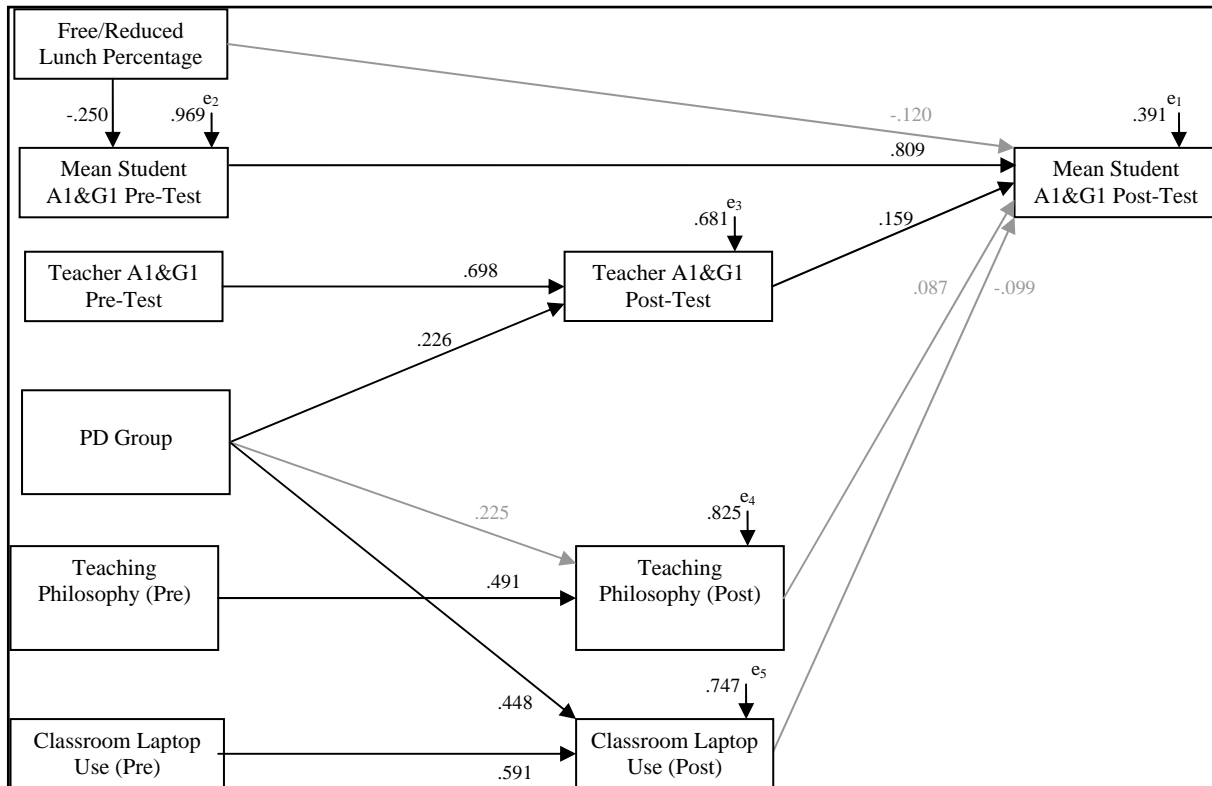


Figure 3: Path Diagram for Total Mathematics Scores - Spring 2006 Sample

Figure 3 confirms several of the simple correlations reported in Table 10, but it also reveals some of the significant paths to student posttest total mathematics achievement scores. One is the significant relationship between group and teacher posttest content knowledge scores and student posttest achievement. Group was also related to post-intervention laptop use levels, but not use levels to student achievement.

Figure 4 displays the same analyses, but breaks the group membership variable into two variables. The group labeled PD Group represents the study sample at Time 3 (i.e., 20+ months continuous involvement group). This represents the teachers in the experimental group who participated continuously in the professional development program over the course of the two-year intervention period. The second group, the FSS group, represents the group of experimental group teachers who not only participated continuously in the professional development program,

but also received high rating scores for their level of use of professional development program activities, and strategies in their classroom instruction.

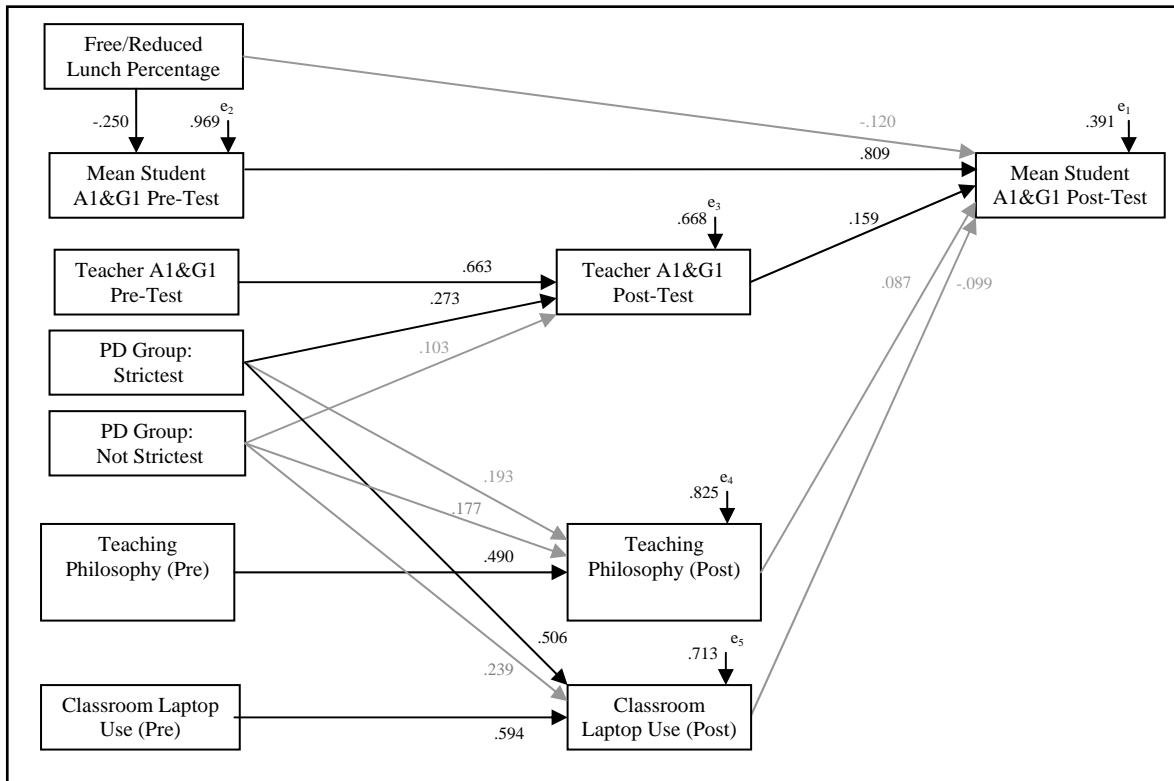


Figure 4: Path Diagram for Total Mathematics Scores - FSS Group Sample

As may be seen in Figure 4, there was a relationship between FSS group membership and post-intervention teacher knowledge and laptop use levels. However, although there was a significant relationship between teacher post-intervention knowledge and student achievement, there was no statistically significant relationship between teacher post-intervention laptop use levels and student achievement.

Figures 5 and 6 describe the path analyses for the A1 subtest. As in the case of total mathematics scores, group was related to teacher posttest scores for the FSS Group sample, but unlike the total score relationships, teacher posttest scores were unrelated to student posttest performance. Neither was laptop use levels related to student performance.

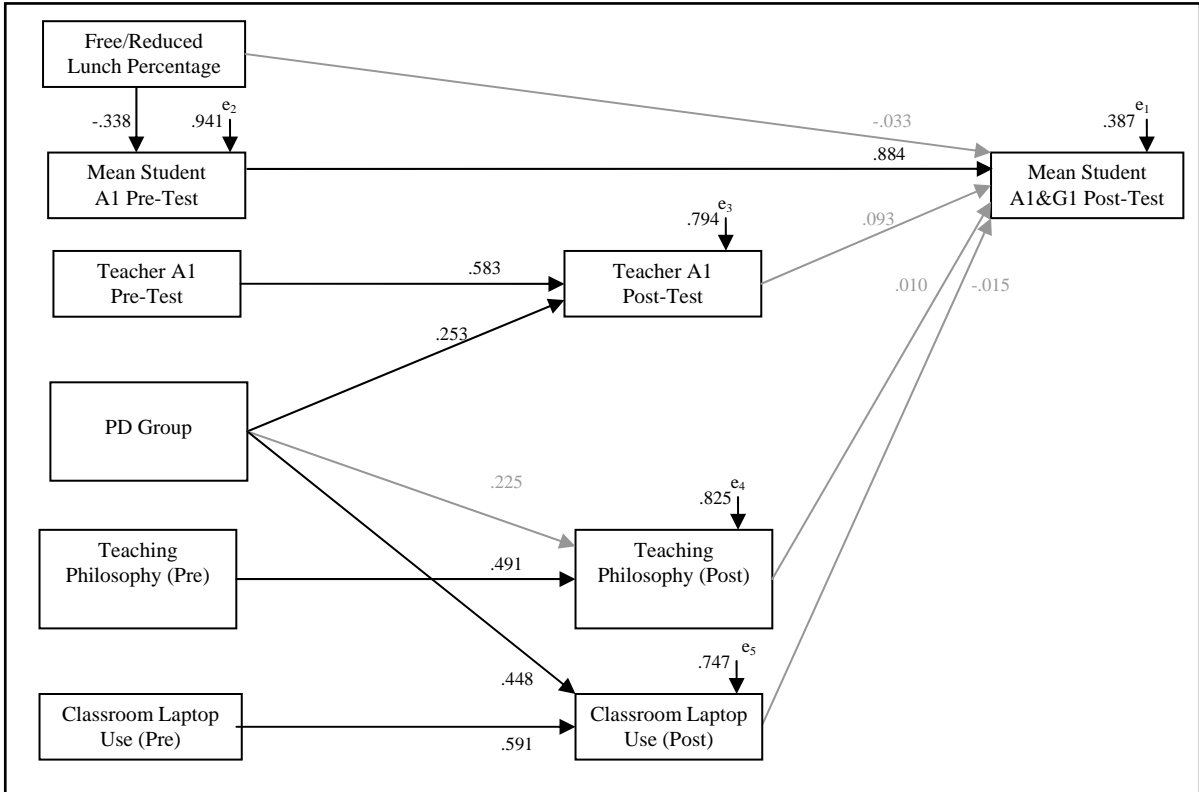


Figure 5: Path Diagram for A1 Subtest Scores - Spring 2006 Sample

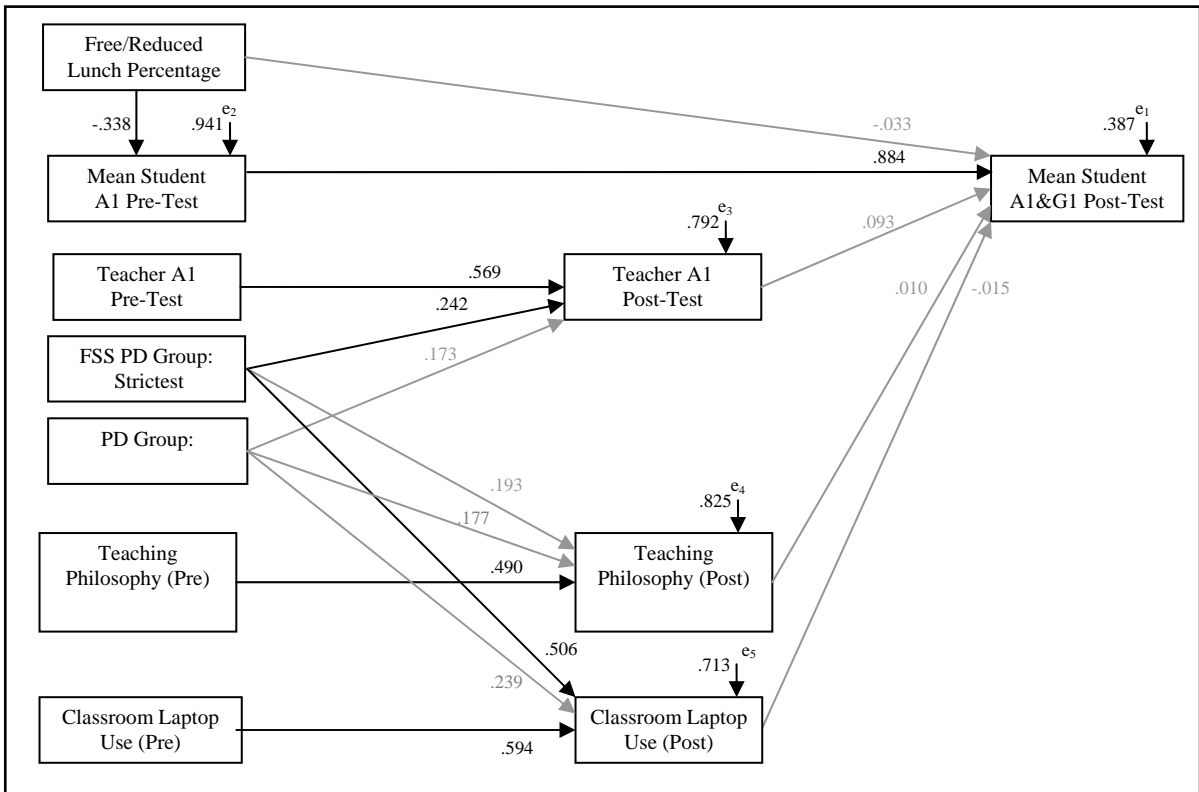


Figure 6: Path Diagram for A1 Subtest Scores - FSS Group Sample

However, some additional relationships were statistically significant in the case of G1/K2 subtest analyses. These become most apparent in Figure 8, where the Spring 2006 groups are broken out into the PD Group and FSS groups. In the case of the FSS group, there are statistically significant relationships between group and teacher posttest content knowledge scores, and between group and post-intervention laptop use levels. And both of these intervening variables, along with post-intervention teaching philosophy are significantly related to student posttest G1 scores.

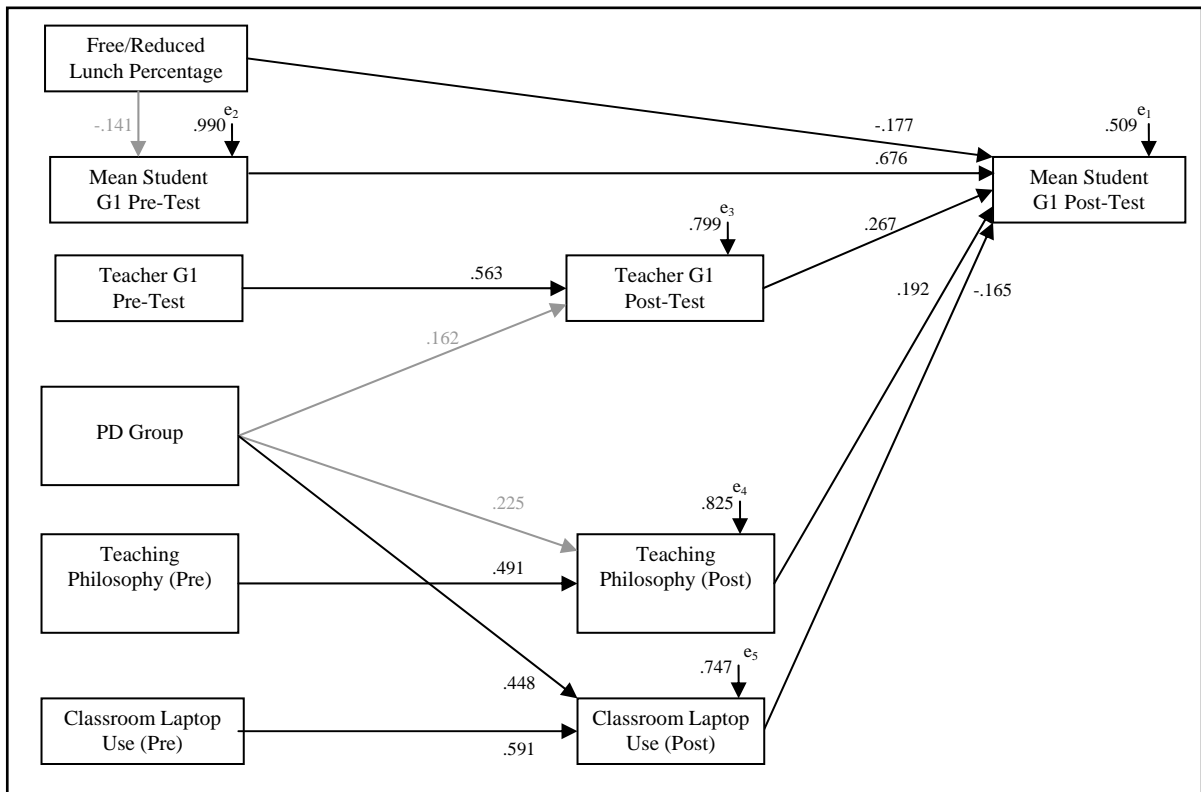


Figure 7: Path Diagram for G1 Subtest Scores - Spring 2006 Sample

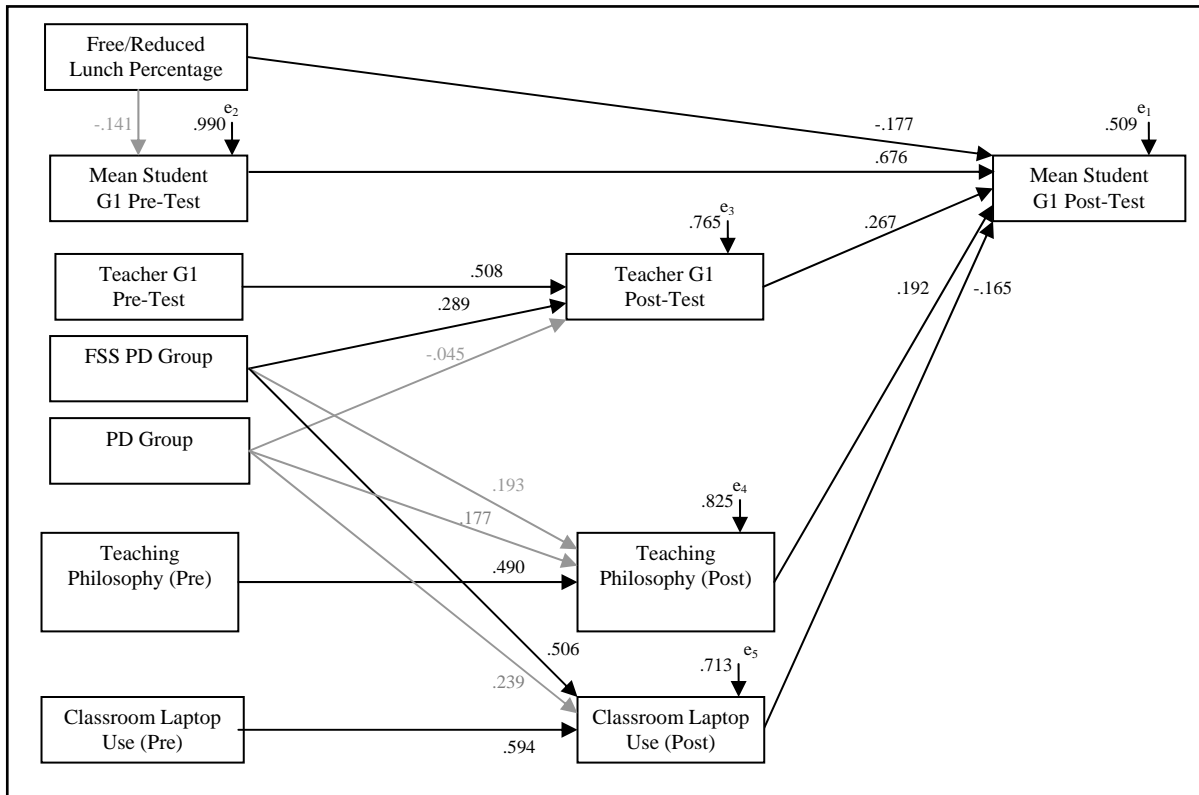


Figure 8: Path Diagram for G1 Subtest Scores - FSS Group Sample

The results from the multiple analyses of the study data indicated mixed findings from the study. In the exploratory analysis, in which the data were analyzed only by student group membership, the experimental group students scored significantly higher on the post-intervention test than the control group students. Further analysis of test results revealed that the experimental group students scored significantly higher on the G1/K2 subtest, but not the A1 subtest.

In the second phase of analysis, using hierarchical linear modeling, group membership was not a statistically significant predictor of student performance. And the causal modeling analyses indicated statistically significant relationships between group membership, teacher knowledge, and student performance for the G1/K2 subtest, but not for the A1 subtests.

Discussion

Why were the results not clearer and more pronounced? Why does it appear the two-year sustained technology-infused professional development program was only moderately successful? There are a myriad of possible explanations, including that the theory is not supported. That is, that the professional development program implemented in this study for

teachers, in reality, cannot be directly linked to improved student performance, at least not for all the content areas examined in this study. As the accumulated evidence suggests, the professional development intervention did result in increased content knowledge for the experimental group of teachers, but this increased knowledge base, and greater use of laptops in teaching mathematics, did not consistently translate into increased student learning. Students' knowledge of *mathematics patterns and relationships* did increase, but not in the areas of *numbers and operations*, the second area of mathematics content which was the focus of the professional development program.

Assuming for the sake of discussion that there is in fact a causal link between sustained professional development and student performance, why was it not clearly established in this RCT study? Implementing randomized controlled trial (RCT) studies, what is referred to as the “Gold Standard” for conducting evidence-based research, faces significant challenges in establishing cause and effect relationships in education (as well as other social science discipline) settings. This study clearly exhibited many of these challenges, and highlighted the tensions between internal and external validity in conducting experimental research. To answer the core research question in this study, several compromises to conducting a RCT were necessary. Participation by schools was voluntary, and while schools were randomly assigned to treatment and control groups, the study experienced several key fidelity of treatment problems. The evidence in this study indicates there was substantial treatment non-compliance, whether they be varying levels of completion of the professional development program, and/or varying levels of implementing intervention activities in classrooms. And students experienced varying levels of intervention implementation across the two-year intervention (e.g., differences in implementation by their 7th and 8th grade teachers). Problems associated with data collection also may have affected the results. Some data points were not collected for various teachers and their students because of local conditions, digital problems, and simple non-compliance. In addition, all student assessments were locally administered (i.e., by classroom teachers) and monitored. The timing of the assessments may also have affected the results. Tests were administered within a two-week period of time (to mirror existing statewide practices for State learning standards assessments) at appointed times over the course of the two-year intervention. However, when teachers implemented the intervention content and instructional strategies varied by school conditions, local curriculum decisions, and teacher professional judgment, which in

turn resulted in considerable variance between the time of classroom implementation and student exposure and assessment of learning. These significant threats to the internal validity of this study may have had substantial impacts on the study results as well as the generalizability of the study findings.

In addition to the fidelity of treatment problems, other factors may be hypothesized as explanations of the mixed results. For one, the assessment results at the end of the 8th grade may reflect differences in the focus of the instruction in the 7th and 8th grades. In most of Maine schools, 7th grade content and instruction focuses more on A1 content, and 8th grade on G1/K2 content. This may explain the mixed results. Secondly, teaching emphasis may have varied. Even though teacher post-intervention content knowledge was found to be related to group membership, teachers may have emphasized more procedural knowledge in their A1 instruction. There was some anecdotal information to this effect, but in the absence of multiple classroom observations over the two years, a strategy that was cost prohibitive, this hypothesis could not be tested. A third possible hypothesis may be that there was a mismatch between the instruction and assessment instrument. G1 content, *Patterns and Relationships*, was assessed, in part, by asking the student to demonstrate their learning using a digital interactive mathematics applet, the same strategy that was used by teachers in teaching the mathematics content, whereas the A1 content, *Numbers and Operations*, was assessed using a traditional paper and pencil test.

Returning to the overall findings from this study, a central question still remains, why were the results not more robust? One possible reason is that there was insufficient time to measure the effects. Possibly the changes in student performance will not be surfaced until later years. And, as just discussed above, the types of student assessment measures may have contributed to the results. But in addition to these possible explanations of the results, as well as others that may be hypothesized, one critical reason for the lackluster results may have been the tensions caused by attempting to conduct rigorous research in education field sites.

Summary

The core question that was explored in this study was:

Does a technology-infused professional development program for teachers in rural schools result in significant improvements in grade 8 students' mathematics achievement?

The results from this study suggest that the answer to this question is a qualified “yes”. That is to say, if a teacher actively participated in the intervention activities for 20 months or more,

increased their own content knowledge, and implemented classroom technology use practices, then student achievement may improve. But student achievement may not increase in all areas, and the relationships between change in laptop use levels and student achievement was weak in this study.

This study also demonstrates many of the issues and obstacles researchers face in balancing internal and external validity factors in attempting to conduct more rigorous scientifically-based research in education field sites. Because schools are not in the business of conducting “gold standard” research, but in the business of educating youth, researchers must make concessions in the research design and execution of a study to accommodate the realities of schooling. But in so doing, the researcher compromises their ability to establish causality.

Given these realities and dilemma does not mean education researchers should abandon attempts to conduct more rigorous studies of education interventions. As Grable and Vasra remark in reflecting on their own work:

“Can high-quality research take place in schools? Absolutely. Can such research inform best practices and guide educational policy? Certainly. Can this research, in all cases, reflect the types of medical models that inform the new federal guidelines for educational inquiry? Probably not.”

Thus, the goal should continue to be to strive to conduct more high quality, rigorous studies in many education areas, including the impacts of technology on student learning.

References

- Ball, D., & Cohen, D. (1999). Developing Practice, Developing Practitioners: Toward a Practice-based Theory of Professional Education. In The Heart of the Matter. In L. Darling-Hammond & L. Sykes (Eds.), *Teaching as the Learning Profession*. San Francisco: Jossey-Bass.
- Ball, D. L., & Bass, H. (2000). Interweaving content and pedagogy in teaching and learning to teach: Knowing and using mathematics. In J. Boaler (Ed.), *Multiple perspectives on the teaching and learning of mathematics* (pp. 83 -104). Westport, CT: Ablex.
- Becker, H. J., Ravitz, J. L., & Wong, Y. (1999). *Teacher and teacher-directed student use of computers and software* (Teaching, Learning, and Computing: 1998 National Survey Report #3). Irvine, CA: Center for Research on Information Technology and Organizations, University of California-Irvine.
- Borasi, R., & Fonzi, J. (2002, August 9). *Professional Development That Supports School Mathematics Reform*. Retrieved 25 Nov. 2002, from the World Wide Web: <http://www.nsf.gov/pubs/2002/nsf02084/start.htm>.
- Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Bryk, A. S., Raudenbush, S. W., & Congdon, R. (1996). *Hierarchical Linear and Nonlinear Modeling with the HLM/2 and HLM/3 Programs*. Chicago: Scientific Software International, Inc.
- Burrill, G., Allison, J., Breaux, G., Kastberg, S., Leatham, K., & Sanchez, W. (2002). *Handheld Graphing Technology in Secondary Mathematics: Research Findings and Implications for Classroom Practice*. Dallas, TX: Texas Instruments.
- Darling-Hammond, L., & McLaughlin, M. W. (1995). Policies That Support Professional Development in an Era of Reform. *Phi Delta Kappan*, 76(8), 597-604.
- Goldstein, H. (1995). *Multilevel Statistical Models*. London: Edward Arnold.
- Grable, A.S., & Vasra, E.S. (2006), Evidence-based education; Postcards from the Edge. *Journal of Technology and Teacher Education*, 14(3), 623-632
- Grant, C. M. (1998). *Bringing Technology Tools in Line with Elementary Mathematics Curricula* (2). Cambridge: TERC.
- Heid, K. M., & Blume, G. W. (Eds.). (2003). *Research on Technology in the Teaching and Learning of Mathematics: Syntheses and Perspectives*. Greenwich, CT: Information Age Publishing Inc.
- Kleiman, G. (2000). Myths and Realities about Technology in K-12 Schools. In D. Gorden (Ed.), *The Digital Classroom* (pp. 7 -15). Cambridge, MA: Harvard Education Letter.
- Loucks-Horsley, S., Hewson, P. W., Love, N., & Stiles, K. E. (1998). *Designing Professional Development for Science and Mathematics*. Thousand Oaks, CA: Corwin Press, Inc.

The Maine Learning Technology Initiative: Teacher, Student, and School Perspectives Mid-Year Evaluation Report. (2003). Gorham, ME: Maine Education Policy Research Institute, Center for Education Policy, Applied Research, & Evaluation, University of Southern Maine.

Meeting the Highly Qualified Teachers Challenge: The Secretary's Second Annual Report on Teacher Quality. (2003). Washington, D.C.: U.S. Department of Education, Office of Policy Planning and Innovation.

National Foundation for the Improvement of Education. (1996). *Teachers Take Charge of Their Learning: Transforming Professional Development for Student Success.* Washington, DC: NFIE.

National Staff Development Council. (2001). *E-Learning for Educators: Implementing the Standards for Staff Development.* Retrieved 3 Apr. 2003, from the World Wide Web: <http://www.nsd.org/e-learning.pdf>

Sandholtz, J. H., Ringstaff, C., & Dwyer, D. C. (1997). *Teaching with Technology.* New York: Teachers College Press.

Schmitz, C.C., & Parsons, B.A. (1999). *Everything You Wanted to Know About Logic Models But Were Afraid to Ask.* Battle Creek, MI. W.K. Kellogg Foundation.

Sparks, D., & Hirsh, S. (1997). *A New Vision for Staff Development.* Alexandria, VA: Association for Supervision and Curriculum Development

Treacy, B., Kleiman, G., & Peterson, K. (2002). Elements of Successful Online Professional Development Programs. *Learning & Leading with Technology*, 30(1.).

U.S. Department of Education (2003). Scientifically base evaluation methods. *Federal Register* 68 (213).

Wenglinsky, H. (1998). *Does it compute? The relationship between educational technology and student achievement in mathematics* (Policy Information Report). Princeton, NJ: Educational Testing Service.