

Concerns, Considerations, and New Ideas for Data Collection and Research in Educational Technology Studies

Damian Bebell
Laura M. O'Dwyer
Michael Russell
Tom Hoffmann
Boston College

Abstract

In the following pages, we examine some common methodological challenges in educational technology research and highlight new data collection approaches using examples from the literature and our own work. Given that surveys and questionnaires remain widespread and dominant tools across nearly all studies of educational technology, we first discuss the background and limitations of how researchers have traditionally used surveys to define and measure technology use (as well as other variables and outcomes). Through this discussion, we introduce our own work with a visual analog “sliding” scale as an example of a new approach to survey design and data collection that capitalizes on the technology resources increasingly available in schools. Next, we highlight other challenges and opportunities inherent in the study of educational technology, including the potential for computer adaptive surveying, and discuss the critical importance of aligning outcome measures with the technological innovation, concerns with computer-based versus paper-based measures of achievement, and the need to consider the hierarchical structure of educational data in the analysis of data for evaluating the impact of technology interventions. (Keywords: research methodology, survey design, measurement, educational technology research)

This paper examines some common methodological issues facing educational technology research and provides suggestions for new data collection approaches using examples from the literature and the authors' own experience. Given that surveys and questionnaires remain widespread and dominant tools across nearly all studies of educational technology, we first discuss the background and limitations of how researchers have traditionally used surveys to define and measure technology use (as well as other variables and outcomes). Through this discussion, we introduce our own approaches and tools that we have used in recent studies that capitalize on the technology resources increasingly available in schools. Next, we highlight

other challenges and opportunities inherent in the study of educational technology, including the potential for computer adaptive surveying, and discuss the critical importance of selecting and aligning outcome measures with the technological innovation, concerns with computer-based versus paper-based measures of achievement, and the need to consider the hierarchical structure of educational data in the analysis of data for evaluating the impact of technology interventions.

The integration of computer technologies into U.S. classrooms over the past quarter century has arguably led to a widespread shift in the U.S. K–12 educational landscape. Believing that increased use of computers will lead to improved teaching and learning, greater efficiency, and the development of important skills among students, educational leaders and policy makers have made multibillion dollar investments in educational technologies. With these investments, the national ratio of students to computers has dropped from 125:1 in 1983 to 4:1 in 2006 (U.S. Census Bureau, 2006). In addition, between 1997 and 2003, the percentage of U.S. classrooms connected to the Internet grew from 27% to 93%. In 1997, 50% of schools used a dialup connection to connect to the Internet, and only 45% had a dedicated high-speed Internet line. By 2003, less than 5% of schools were still using dialup connections, whereas 95% reported having broadband access. In a relatively short time period, computer-based technologies have become commonplace across all levels of the U.S. educational system. Given these substantial investments in educational technology, it is not surprising that there have been calls over the past decade for empirical, research-based evidence that these massive investments are affecting the education and lives of teachers and students (Cuban, 2006; McNabb, Hawkes, & Rouk, 1999; Roblyer & Knezek, 2003; Weston & Bain, 2009).

Several advances in computer-based technologies converged in the mid-1990s to greatly increase the capacity for computer-based technology to support teaching. As increased access and more powerful computer-based technologies entered U.S. classrooms, the variety of ways and the degree to which teachers and students applied these new technologies increased exponentially. Whereas in the early days of educational technology, integration instructional uses of computers had been limited to word processing, skills software, and computer programming, teachers were now able to perform multimedia presentations and computer-based simulations. With the introduction of the Internet into the classroom, teachers were also able to incorporate activities that tapped the resources of the World Wide Web. Outside of class time, software for recordkeeping, grading, and test development provided teachers with new ways of using computers to support their teaching. In addition, the Internet allowed teachers access to additional resources when planning lessons and activities (Becker, 1999; Zucker & Hug, 2008), and allowed teachers to use email to communicate with their colleagues, administrative leaders, students, and parents (Bebell & Kay, 2009; Lerman, 1998).

Following the rise of educational technology resources, hundreds of studies have sought to examine instructional uses of technology across a wide variety of educational settings. Despite the large number of studies, many researchers and decision makers have found past and current research efforts unsatisfactory. Specifically, criticisms of educational technology research have focused on both the lack of guiding theory as well as the failure to provide adequate empirical evidence on many salient outcome measures (Roblyer & Knezek, 2003; Strudler, 2003; Weston & Bain, 2010). For example, in Roblyer and Knezek's (2003) call for a national educational technology research agenda, they declare that the next generation of scholarship "must be more comprehensive and informative about the methods and materials used, conditions under which studies take place, data sources and instruments, and subjects being studied; and they must emphasize coherence between their methods, findings, and conclusions" (p. 69).

Although there has been more examination and discussion about general research shortcomings, many critics and authors have examined and highlighted specific weaknesses across the published literature. Baker and Herman (2000); Waxman, Lin, and Michko (2003); Goldberg, Russell, and Cook (2003); and O'Dwyer, Russell, Bebell, and Tucker-Seeley (2005, 2008) have all suggested that many educational technology studies suffer from a variety of specific methodological shortcomings. Among other deficits, past reviews of educational technology research found it was often limited by the way student and teacher technology use was measured, a poor selection/alignment of measurement tools, and the failure to account for the hierarchical nature of data collected from teachers and students in schools (Baker & Herman, 2000; Goldberg, Russell, & Cook, 2003; O'Dwyer, Russell, Bebell, and Tucker-Seeley, 2005, 2008; Waxman, Lin, & Michko, 2003).

The collective weaknesses of educational technology research has created a challenging situation for educational leaders and policy makers who must use flawed or limited research evidence to make policy and funding decisions. Even today, little empirical research exists to support many of the most cited claims on the effects of educational technology.¹ For example, despite a generation of students being educated with technology, there has yet to be a definitive study that examines the causal impacts of computer use in school on standardized measures of student achievement. It is a growing problem that, as an educational community, research and evaluation efforts have not adequately elucidated the short- and long-term effects of technology use in the classroom. This situation forces decision makers to rely on weak sources of evidence, if any at all, when allocating budgets and shaping future policy around educational technology.

¹ Educational technology research is often divided into two broad categories: (a) research that focuses on effects with technology in the classroom and (b) research that focuses on the effects of technology integrated into the classroom and teacher practices (Salomon, Perkins, & Globerson, 1991). Although not mutually exclusive, this categorization of research can be illuminating. Generally, research concerning the "effects with" technology focuses on the underlying evolution of the learning process with the introduction of technology. On the other hand, research concerning the "effect of" technology seeks to measure (via outcomes testing) the impacts of technology as an efficiency tool rather than focusing on the underlying processes. The current paper concentrated more on the latter category, the "effects of" technology.

Not surprisingly, in today's zeitgeist of educational accountability, the call for empirical, research-based evidence that these massive investments are affecting the lives of teachers and students has only intensified. It is our hope that the current paper will serve to further illuminate a small number of methodological limitations and concerns that affect much of the educational technology research literature, and will provide real-world examples from our own efforts on promising approaches and techniques to improve future inquiries.

Defining and Measuring Technology Use with Surveys

Since the earliest adoption of computer-based technology resources in education, there has been a desire to collect empirical evidence on the impact of technology on student achievement and other outcome variables. The impacts on learning, however, must be placed in the context of technology use. Before the impact of technology integration can be studied, there must be solid empirical evidence of how teachers and students are using technology. As such, sound research on the impact of technology integration is predicated on the development and application of valid and reliable measures of technology use.

To measure technology use appropriately (as well as any other variable or indicator), researchers must invest time and effort to develop instruments that are both reliable and valid for the inferences that are made. Whether collected via paper or computer, survey instruments remain one of the most widely employed tools for measuring program indicators. However, the development of survey items poses particular challenges for research that focuses on new and novel uses of technology. Because the ways a given technology tool is used can vary widely among teachers, students, and classrooms, the survey developer must consider a large number of potential uses for a given technology-based tool to fully evaluate its effectiveness.

For decades, paper-and-pencil administrations of questionnaires or survey instruments dominated research and evaluation efforts, but in recent years an increasing number of researchers are finding distinct advantages to using Internet-based tools for collecting their data (Bebell & Kay, 2010; Shapley, 2008; Silvernail, 2008; Weston & Bain, 2010). Web-based surveys are particularly advantageous in settings where technology is easily accessible, as is increasingly the case in schools. In addition, data collected from computer-based surveys can be accessed easily and analyzed nearly instantly, streamlining the entire data collection process. However, the constraints and limitations of paper-based surveys have been rarely improved upon in their evolution to computer-based administration; typically, technology-related surveys fail to capitalize on the affordances of technology-based data collection. In the extended example below, we use the example of measuring teachers' use of technology to (a) explore how teachers' use of educational technology has been traditionally defined and measured in the educational

technology literature, (b) demonstrate the limitations and considerations when quantifying the frequency of technology use with a traditional survey design, and (c) introduce our visual analog “sliding” scale, which capitalizes on the availability of technology resources in schools to improve the accuracy and validity of traditional survey design.

Defining Technology Use

A historical review of the literature on educational technology reveals that the definition of *technology use* varies widely across research studies. The first large-scale investigation of modern educational technology occurred in 1986 when Congress asked the federal Office of Technology Assessment (OTA) to compile an assessment of technology use in U.S. schools. Through a series of reports, OTA (1988, 1989, 1995) documented national patterns of technology integration and use. Ten years later, Congress requested that OTA “revisit the issue of teachers and technology in K–12 schools in depth” (OTA, 1995, p. 5). In a 1995 OTA report, the authors noted that previous research on teachers’ use of technology employed different definitions of what constituted technology use. In turn, these different definitions led to confusing and sometimes contradictory findings regarding teachers’ use of technology. By way of another example, a 1992 International Association for the Evaluation of Educational Achievement (IEA) survey defined a “computer-using teacher” as a teacher who “sometimes” used computers with students. A year later, Becker (1994) employed a more explicit definition of a computer-using teacher for which at least 90% of the teachers’ students were required to use a computer in their class in some way during the year. Thus, the IEA defined use of technology in terms of the teachers’ use for instructional delivery, whereas Becker defined use in terms of the students’ use of technology during class time. It’s no surprise that these two different definitions of a *computer-using teacher* yielded different impressions of the extent of technology use. In 1992, the IEA study classified 75% of U.S. teachers as computer-using teachers, whereas Becker’s criteria yielded about one third of that (approximately 25 %) (OTA, 1995). This confusion and inconsistency led OTA to remark: “Thus, the percentage of teachers classified as computer-using teachers is quite variable and becomes smaller as definitions of use become more stringent” (p. 103).

In the decade(s) since these original research efforts, teachers’ use of technology has increased in complexity as technology has become more advanced, varied, and pervasive in schools, further complicating researcher efforts to define and measure “use.” Too often, however, studies focus on technology access instead of measuring the myriad ways that technology is being used. Such research assumes that teachers’ and students’ access to technology is an adequate proxy for the use of technology. For example, Angrist and Lavy (2002) sought to examine the effects of educational technology on student achievement using Israeli standardized test data. In their study, the authors

did not measure student or teacher practices with technology, but compared levels of academic achievement among students classified as receiving instruction in either high- or low-technology environments. In other words, the research had no measures of actual technology use, but instead classified students based on their access to technology. Although access to technology has been shown to be an important predictor of technology use (Bebell, Russell, & O'Dwyer, 2004; Ravitz, Wong, & Becker, 1999), a wide variety of studies conducted in educational environments where technology access is robust, yet use is not, suggest that the assumption is inadequate for research that is used to inform important educational and policy decisions around educational technology (Bebell, Russell, & O'Dwyer, 2004; Weston & Bain, 2010). Clearly, measuring access to computers is a poor substitute for the measurement of actual use in empirical research, a point that is further highlighted when readers learn that Angrist and Lavy's well publicized 2002 study defined and classified settings where 10 students shared a single computer (i.e., a 10:1 ratio) as the "high-access schools."

Today, several researchers and organizations have developed their own definitions and measures of technology use to examine the extent of technology use and to assess the impact of technology use on teaching and learning. Frequently these instruments collect information on a variety of types of technology use and then collapse the data into a single generic "technology use" variable. Unfortunately, the amalgamated measure may be inadequate both for understanding the extent to which technology is being used by teachers and students, and for assessing the impact of technology on learning outcomes (Bebell, Russell, & O'Dwyer, 2004). Ultimately, decision makers who rely on different measures of technology use will likely come to different conclusions about the prevalence of use and its relationship with student learning outcomes. For example, some may interpret one measure of teachers' technology use solely as teachers' use of technology for delivering instruction, whereas others may view it as a generic measure of a teacher's collected technology skills and uses. Although defining technology use as a single dimension may simplify analyses, it complicates efforts by researchers and school leaders to provide valid and reliable evidence of how technology is being used and how use might relate to improved educational outcomes.

Recognizing the Variety of Ways Teachers Use Technology

One approach to defining and measuring technology use that we have found effective has concentrated on developing multiple measures that focus on specific ways that teachers use technology. This approach was employed by Mathews (1996) and Becker (1999) in demonstrating the complicated relationship between teachers' adoption and use of technology to support their teaching. Similarly, in our own effort to better define and measure the ways teachers use technology to support teaching and

learning, we examined survey responses from more than 2,500 K–12 public school teachers who participated in the federally funded USEIT Study (Russell, O’Dwyer, Bebell, & Miranda, 2003). Analyzing these results using factor analytic techniques we developed seven distinct scales that measure teachers’ technology use:

- Teachers’ use of technology for class preparation
- Teachers’ professional e-mail use
- Teacher-directed student use of technology during class time
- Teachers’ use of technology for grading
- Teachers’ use of technology for delivering instruction
- Teachers’ use of technology for providing accommodations
- Teacher-directed student use of technology to create products

Analyses that focused on these seven teacher technology use scales revealed that the frequency with which teachers employed technology for each of these purposes varied widely (Bebell, Russell, & O’Dwyer, 2004). For example, teachers’ use of technology for class preparation was strongly negatively skewed (skewness = -1.12), inferring that a majority of surveyed teachers frequently used technology for planning, whereas only a small number of teachers did not. Conversely, the use of technology for delivering instruction was strongly positively skewed (1.09), meaning that the majority of surveyed teachers rarely used technology to deliver instruction, whereas most reported never or only rarely using technology to deliver instruction. Distributions for teacher-directed student use of technology to create products (1.15) and teachers’ use of technology for providing accommodations (1.04) were also positively skewed. Using technology for grading had a weak positive skew (0.60), whereas teacher-directed student use of technology during class time (0.11) was relatively normally distributed. Teachers’ use of e-mail, however, presented a bimodal distribution, with a large percentage of teachers reporting frequent use and a large portion of the sample reporting no use. Interestingly, when these individual scales were combined into a generic “technology use” scale (as is often done with technology use surveys), the distribution closely approximated a normal distribution. Thus, the generic technology use measure obscured all of the unique and divergent patterns observed in the specific technology use scales (Bebell, Russell, & O’Dwyer, 2004).

Clearly, when compared to a single generic measure of technology use, using multiple measures of specific technology use offers a more nuanced understanding of how teachers use technology and how these uses vary among teachers. Research studies that have utilized this multifaceted approach to measuring technology use have revealed many illuminative patterns that were obscured when only general measures of use were employed (Bebell, Russell, & O’Dwyer, 2004; Mathews, 1996; Ravitz, Wong, & Becker,

1999). For example, when we examined teachers' use of technology using a generic measure that compromised a wide variety of types of technology use, it appeared that the frequency with which teachers use technology did not vary noticeably across the number of years they had been in the profession. In other words, teachers who were brand new to the profession appeared to use technology as frequently as teachers who had been in the profession for 11 or more years. However, when distinct individual types of technology use were examined, newer teachers reported higher levels of technology use for preparation and slightly higher levels of use for accommodating students' special needs than did more experienced teachers. Conversely, new teachers reported less frequent use of technology for instructional use and having their students to use technology during class time than their more experienced colleagues (Bebell, Russell, & O'Dwyer, 2004). These examples convey the importance of fully articulating and measuring technology use and how different measures of technology use (even with the same data set) can lead to substantially varied results.

How technology use is defined and measured (if measured at all) plays a substantial, but often overlooked, role in educational technology research. For example, using NAEP data, Wenglinski (1998) employed two measures of technology use in a study on the effects of educational technology on student learning. The first measure focused specifically on use of technology for simulation and higher-order problem solving and found a positive relationship between use and achievement. The second measure employed a broader definition of technology use and found a negative relationship between use and achievement. Thus, depending how one measures use, the relationship between technology use and achievement appeared to differ.

Similarly, O'Dwyer, Russell, Bebell, and Tucker-Seeley (2005) examined the relationship between various measures of computer use and students English/language arts test scores across 55 intact upper elementary classrooms. Their investigation found that, while controlling for both prior achievement and socioeconomic status, students who reported greater frequency using technology in school to edit their papers also exhibited higher total English/language arts test scores and higher writing scores. However, other measures of teachers' and students' use of technology, such as students' use of technology to create presentations and recreational use of technology at home were not associated with increased English/language arts outcome measures. Again, different findings related to how "technology use" was associated with student test performance resulted depending on how the researchers chose to define and measure technology use. These examples typify the complex, and often contradictory, findings that policy makers and educational leaders confront when using educational technology research to guide policy-related technology decisions.

Table 1. Assigning Linear Values to Represent Use

Response Option	Assigned Value
<input type="checkbox"/> Never	0
<input type="checkbox"/> Once or twice a year	1
<input type="checkbox"/> Several times a year	2
<input type="checkbox"/> Once a month	3
<input type="checkbox"/> Several times a month	4
<input type="checkbox"/> Once a week	5
<input type="checkbox"/> Several times a week	6
<input type="checkbox"/> Everyday	7

Four Approaches for Representing the Frequency of Technology Use

Below, we present an extended example of how teachers' technology use is typically measured via a survey instrument, including clear limitations to traditional approaches and recommendations for capitalizing on the affordances provided by technology for improving overall accuracy and validity. Traditionally, surveys present respondents with a set of fixed, close-ended response options from which they must select their response. For example, when measuring the frequency of technology use, teachers may be asked to select from a discrete number of responses for a given item. As an example, the survey question below (adapted from the 2001 USEIT teacher survey) asks a teacher the frequency with which they used a computer to deliver instruction:

During the last school year, how often did you use a computer to deliver instruction to your class?

- Never
- Once or twice a year
- Several times a year
- Once a month
- Several times a month
- Once a week
- Several times a week
- Everyday

(Russell, et al., 2003)

Table 2: Assigning “Real” Values to Represent Use

Response Option	Assigned Value
<input type="checkbox"/> Never	0
<input type="checkbox"/> Once or twice a year	2
<input type="checkbox"/> Several times a year	6
<input type="checkbox"/> Once a month	9
<input type="checkbox"/> Several times a month	27
<input type="checkbox"/> Once a week	36
<input type="checkbox"/> Several times a week	108
<input type="checkbox"/> Everyday	180

In this example, a respondent selects the response option that best represents the frequency with which s/he uses a computer to deliver instruction. To enable the statistical analyses of the results, the researcher must assign a numeric value to each of the potential response options. Using the current example, the number assigned to each response option would correspond linearly with increasingly frequent technology use (for example, Never = 0 to Everyday = 7). This 8-point scale (0–7) differentiates how frequently each teacher uses technology for instruction over the course of a given year. By quantifying the responses numerically, a variety of arithmetic and statistical analyses may be performed.

In measurement theory, a greater number of response options provides greater mathematical differentiation of a given phenomenon, which in this case is the frequency of technology use. However, requiring respondents to select a single response from a long list of options can become tedious and overwhelming. Conversely, using fewer response options provides less differentiation among respondents and less information about the studied phenomenon. As a compromise, survey developers have typically employed 5- to 7-point scales to provide a balance between the detail of measurement and the ease of administration (Dillman, 2000; Nunnally, 1978).

However, this widely employed approach has an important limitation. Using the current example, the response options are assigned using linear one-step values, whereas the original response options describe nonlinear frequencies. Linear one-step values result in an ordinal measurement scale, “where values do not indicate absolute qualities, nor do they indicate the intervals between the numbers are equal” (Kerlinger, 1986, p. 400). From a measurement point of view, the values assigned in the preceding example are actually arbitrary (with the exception of 0, which indicates that a teacher never uses technology). Although this type of scale serves to differentiate degrees of teachers’ technology use, the values used to describe the frequency of use are unrelated to the original scale. Consider the example in which this survey question was administered to a sample of middle school

teachers at the beginning and again near the end of the school year. The average value calculated across all teachers during the first administration was 2.5, indicating that, on average, teachers used technology for instruction between several times a year and once a month. The average value calculated across the teachers during the second administration was 5.1, or about once per week. Arithmetically, it appears that the frequency with which teachers use technology has doubled. This doubling, however, is an artifact of the scale assigned to the response options and does not accurately reflect the actual change in the frequency of use.

Table 2 displays an alternate coding system in which the assigned values for each response option are designed to reflect the actual frequency with which teachers could use technology to deliver instruction over the course of a 180-day school year.

In this example, the same survey question and response options are presented; however, the researcher assigns values to each response choice that represent “real” values. Assuming the school year equals 180 days (or nine months, 36 weeks) the analyst assigns values to each response option that reflects the estimated frequency of use. This approach results in a 180-point scale, where 0 represents a teacher never using technology and 180 represents everyday use of technology. This approach provides easier interpretation and presentation of summary data, because the difference between the numbers actually reflects an equal difference in the amount of attribute measured (Glass & Hopkins, 1996).

In the current example, the resulting survey data takes on qualities of an interval measurement scale, whereby “equal differences in the numbers correspond to equal differences in the amounts the attributes measure” (Glass & Hopkins, 1996, p. 8). In other words, rather than the 8-step scale presented in the first example, the 181-step scale offers a clearer and more tangible interpretation of teachers’ technology use. The number of times a teacher may use technology can occur at any interval on a scale between 0 and 180; however, in the current example, teachers responding to the item were still provided with only eight discrete response options in the original survey question. The small number of response options typically employed in survey research forces survey respondents to choose a response-option answer that best approximates their situation. For example, a teacher may use technology somewhat more than once a week but not quite several times per week. Faced with inadequate response options, the teacher must choose between the two options. In this scenario, the survey respondent is forced to choose one of the two available options, both of which yield imprecise, and ultimately inaccurate, data. If the teacher selects both options, the analyst typically must discard the data or be forced to subjectively assign a value to the response. Thus, whenever a survey uses limited response options to represent the frequency of an activity, the collected data may particularly suffer from measurement error if it provides only limited numbers of response choices.

Recognizing the measurement limitations of limited response options in traditional survey design, as well as the increasing presence of technology in educational settings, we have experimented with ways of improving the accuracy of our data collection efforts through the use of new technology-enabled tools to improve traditional survey data collection. Specifically, across our recent studies, we have developed and applied an online survey presentation method where survey items are presented with continuous scales that allow the respondent to select a value that accurately reflects their technology use rather than relying on a limited number of fixed, closed-ended response options (Bebell & Russell, 2006; Bebell, O'Dwyer, Russell, & Hoffmann, 2007; Tucker-Seeley, 2008). Through the use of Macromedia Flash visual analog scale, survey respondents are presented with a full, but not overwhelming, range of response options. This advancement in data collection technology allows the same survey item to be measured using a ratio scale that presents the entire range of potential use (with every available increment present) to teachers. In the following example, teachers are presented the technology use survey question with the visual analog scale in Figure 1.

To complete the survey item, each respondent uses a mouse/trackpad to select the response on the sliding scale. Although the interactive nature of the visual analog scale is challenging to demonstrate on paper, the program is designed to help respondents quickly and accurately place themselves on the scale. In the current example, the teachers' response is displayed for them (in red) under the heading "approximate number of times per year." As a survey respondent moves the sliding scale across the response options on the horizontal line, the "approximate number of times per year" field displays their response in real time. Thus, a teacher can move the slider to any response option between 0 (never) and 180 (daily). In addition, the descriptions above the horizontal slider provide some familiar parameters for teachers so they can quickly select the appropriate response. By solving many of the limitations of traditional categorical survey response options, the visual analog scale provides one example of how digital technologies can be applied to improve traditional data collection efforts in educational technology research.

The Potential for Computer Adaptive Surveying

The application of new technologies in survey research and other data collection efforts can provide many possibilities for improving the quality of educational technology research. Similarly, computer adaptive surveying (CAS) represents the state of the art in development of survey design. In contrast to the current Web-based surveys used to collect data, which present all respondents with a limited set of items in a linear manner, CAS tailors the presentation of survey questions to respondents based on prior item responses. This type of surveying builds upon the theory and

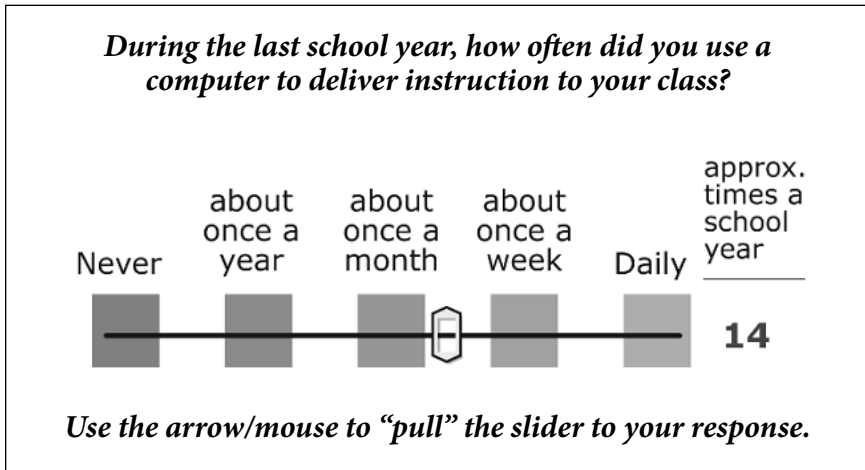


Figure 1. Flash visual analog “sliding” scale.

design of computer adaptive achievement tests, which have been found to be more efficient and accurate than comparable paper-based tests for providing cognitive ability estimates (Wainer, 1990). Similarly, a CAS can tailor the survey questions presented to a given student or teacher to probe the specific details of a general phenomenon.

Take, for example, a recent study we conducted examining how middle school teachers and students use computers in a multischool one-to-one (1:1) laptop program (Bebell & Kay, 2009). Past research and theory suggested that teachers and students across the multiple study settings would likely use computers in very different and distinct ways. So a computer adaptive survey enabled our research team to probe the specific ways teachers and students used technology without requiring them to respond to sets of questions that were unrelated to the ways they personally used computers. Thus, if a student reported that she had never used a computer in mathematics class, the survey automatically skipped ahead to other questions in other subject areas. However, if a student reported that he had used a computer in mathematics class, he would be presented with a series of more detailed and nuanced questions regarding this particular type of technology use (including their frequency of using spreadsheets, modeling functions, etc.).

In another recent pilot study, researchers collaborating with the New Hampshire Department of Education created a Web-based school capacity index to estimate the extent to which a given school will have the technological capacity to administer standardized assessments via computer (Fedorchak, 2008). For this instrument, respondents are first asked about the location and/or type of computers that can be used for testing (labs/media centers, classroom computers, individual student laptops, and/or shared laptops). Then, depending on the answers to the initial question sets, a series of subsequent questions are presented to each respondent that

are uniquely nuanced and specific to their original descriptions of technology access.

Through such adaptive surveys, a more complete and accurate descriptive understanding of a given phenomenon can be acquired. Moreover, due to the adaptive nature of the survey, students and teachers are no longer presented with sets of unrelated survey items, thus decreasing time required to complete the survey, decreasing fatigue, and increasing the accuracy of information collected. Although the use of computer adaptive testing has revolutionized the speed and accuracy of such widespread international assessments as the Graduate Record Exam (GRE) and the Graduate Management Admission Test (GMAT), few examples outside of psychological surveys employ such an approach for data collection in research and evaluation studies. Given the scarcity of time for data collection in most educational settings and the wide variety of technology uses and applications often under review, CAS presents a particularly promising direction for educational technology research.

Use and Alignment of Standardized Tests as Outcomes Measures

Thus far, this paper has largely focused on the data collection aspects of educational technology research and ways that educational technology may be improved by the use of surveys that would improve data collection. However, survey data collection and measurement represent only one aspect of the overall research or evaluation undertaking. In many instances, data collected through surveys is not alone sufficient to address the outcomes of an educational technology study. More typically, studies of educational technology seek to document the impacts of educational technology on measures of student learning, such as classroom or standardized tests.

To adequately estimate any potential impact of educational technology on student learning, all measures of educational outcomes must first be carefully defined and aligned with the specific uses and intended effects of a given educational technology. In other words, when examining the impact of educational technology on student learning, it is critical that the outcome measures assess the types of learning that may occur as a result of technology use and that those measures are sensitive enough to detect potential changes in learning that may occur. By federal law, all states currently administer grade-level tests to students in grades 3–8 in addition to state assessments across different high school grade levels and/or end-of-course tests for high school students. So, for many observers of educational technology programs, such state test results provide easily accessible educational outcomes. However, because most standardized tests attempt to measure a domain broadly, standardized test scores often do not provide measures that are aligned with the learning that may occur when technology is used to develop specific skills or knowledge. Given that the intent and purpose of most state tests is to broadly sample test content across the state standards, such tests often fail to

provide valid measures of the types of learning that may likely occur when students and/or their teachers use computers.

For example, imagine a pilot setting where computers were used extensively in mathematics classes to develop students' understanding of graphing and spatial relationships but infrequently for other concepts. Although the state mathematics test may contain some items relating specifically to graphing and spatial relationships, it is likely that these two concepts will only represent a small portion of the assessment and would be tested using only a very limited number of items, if at all. As a result, researchers using the total math test score would be unlikely to observe any effects of computer use on these two concepts. However, our own research suggests that it may be possible to focus on those subsets of test items that specifically relate to the concepts in question.

In a recent study of the relationship between students' use of technology and their mathematics achievement, we used the state's mandatory Massachusetts Comprehensive Assessment System (MCAS) test as our primary outcome measure (O'Dwyer, Russell, Bebell, & Tucker-Seeley, 2005, 2008). Recognizing that the MCAS mathematics test assesses several different mathematics subdomains, we examined students' overall mathematics test score as well as their performance within five specific subdomains comprising the test:

- Number sense and operations
- Patterns, relationships, and algebra
- Geometry
- Measurement
- Data analysis, statistics, and probability

Through these analyses, we discovered that the statistical models we constructed for each subdomain accounted only for a relatively small percent of the total variance that was observed across students' test scores. Specifically, the largest percentage of total variance explained by any of our models occurred for the total test score (16%), whereas each subdomain scores accounted for even less variance, ranging from 5% to 12% (O'Dwyer, Russell, Bebell, & Tucker-Seeley, 2008). In part, the low amount of variance accounted for by these models likely resulted from the relatively poor reliability of the subtest scores on the MCAS, as each subdomain was composed of a relatively small number of test items; the subdomain measures on the mathematics portion of the fourth grade MCAS test had lower reliability estimates than the test in total. Specifically, the Cronbach's alpha for the fourth grade MCAS total mathematics score was high at 0.86, but the reliabilities of the subdomain scores were generally lower, particularly for those subdomains measured with the fewest number of items. For example, the reliability estimate for data analysis, statistics, and probability subdomain measured with seven items was 0.32, and the reliability for the measurement subdomain measured with only four items was 0.41. The magnitudes

of the reliabilities have important implications for this research because that unreliability in the outcome variable likely makes it more difficult to isolate statistically significant relationships. In other words, despite our best efforts to examine specific types of impacts of educational technology—using subsets of the total state assessment, we observed that serious psychometric limitations could result from insufficient numbers of test items in any one particular area.

Thus, there are many challenges and considerations when measuring student achievement using state assessment scores, even when subdomains of the total test are aligned with practices. Rather than employing state test results, one alternate strategy is to develop customized tests that contain a larger number of items specifically aligned to the types of learning that the educational technology is designed to affect. Although it can be difficult to convince teachers and/or schools to administer an additional test, well-developed aligned assessments will likely result in more reliable scores and provide increased validity for inferences about the impacts of technology use on these concepts.

Paper versus Computer-Based Assessments

In addition to aligning achievement measures with the knowledge and skills students are believed to develop through the use of a given technology, it is also important to align the method used to measure student learning with the methods students are accustomed to using to develop and demonstrate their learning in the classroom. As an example, a series of experimental studies by Russell and colleagues provides evidence that most states' paper-based standardized achievement tests are likely to underestimate the performance of students who are accustomed to working with technology simply because they do not allow students to use these technologies when being tested (Bebell & Kay, 2009; Russell, 1999; Russell & Haney, 1997; Russell & Plati, 2001). Through a series of randomized experiments, Russell and his colleagues provide empirical evidence that students who are accustomed to writing with computers in the classroom perform between 0.4 and 1.1 standard deviations higher when they are allowed to use a computer to perform tests that require them to compose written responses (Russell, 1999; Russell & Haney, 1997; Russell & Plati, 2001).

Other studies replicate similar results, further demonstrating the importance of aligning the mode of measurement with the tools students use (Horkay, Bennett, Allen, Kaplan, & Yan, 2006). One of our more recent studies focused on the impact of a pilot 1:1 laptop program across five middle schools on a variety of outcome measures, including students' writing skills (Bebell & Kay, 2010). Following two years of participation in technology-rich classrooms, seventh grade students were randomly selected to complete an extended writing exercise using either their laptops or the traditional paper/pencil mode espoused by the state. Students in the "laptop" environment submitted a total of 388 essays, and 141 other students submitted

essays on paper were collected on paper before a team of trained readers transcribed and scored them. The results of this study found that students who used their laptops wrote longer essays (388 words compared to 302) and that these essays received higher scores than students responding to the same prompt and assessment using traditional paper and pencil (Bebell & Kay, 2009). These differences were found to be statistically significant, even after controlling for achievement using students' writing scores on the state test that was completed in a traditional testing environment. These results highlight the importance of the mode of measurement in studies looking to explore the impact of educational technology. Specifically, the mode of administration effect suggests that researchers studying the impact of educational technology are particularly at risk for underestimating the ability of technology-savvy students when they rely on paper-based assessment instruments as their outcome measures.

The Hierarchical Nature of Educational Data

A final and related challenge to evaluating the effects of educational technology programs on teaching and learning is the inherent hierarchical nature of data collected from teachers and students in schools. Researchers, evaluators, and school leaders frequently overlook the clustering of students within teachers [classes?] and teachers within schools as they evaluate the impact of technology programs. As a consequence, many studies of educational technology fail to properly account, both statistically and substantively, for the organizational characteristics and processes that mediate and moderate the relationship between technology use and student outcomes. At each level in an educational system's hierarchy, events take place and decisions are made that potentially impede or assist the events that occur at the next level. For example, decisions made at the district or school levels may have profound effects on the technology resources available for teaching and learning in the classroom. As such, researchers and evaluators of educational technology initiatives must consider the statistical and substantive implications of the inherent nesting of technology-related behaviors and practices within the school context.

From a statistical point of view, researchers have become increasingly aware of the problems associated with examining educational data using traditional analyses such as ordinary least-squares regression analysis or analysis of variance. Because educational systems are typically organized in a hierarchical fashion, with students nested in classrooms, classrooms nested in schools, and schools nested within districts, a hierarchical or multilevel approach to data analysis is often required (Burstein, 1980; Cronbach, 1976; Haney, 1980; Kreft & de Leeuw, 1998; Raudenbush & Bryk, 2002; Robinson, 1950). A hierarchical data analysis approach is well suited for examining the effects of technology initiatives. Regardless of whether the outcome of interest is student achievement, affective behaviors, or teacher practices, this approach has three distinct

advantages over traditional analyses. First, the approach allows for the examination of the relationship between technology use and the outcome variable to vary as a function of classroom, teacher, school, and district characteristics. Second, the approach allows the relationship between technology use and the outcome to vary across schools and permits modeling of the variability in the relationship. Third, differences among students in a classroom and differences among teachers can be explored at the same time, therefore producing a more accurate representation of the ways in which technology use may be related to improved educational outcomes (Goldstein, 1995; Kreft & de Leeuw, 1998; Raudenbush & Bryk, 2002).

To date, only a handful of published studies in educational technology research have applied a hierarchical data analysis approach. For example, using data collected from both teachers and students in 55 intact fourth grade classrooms, O'Dwyer and colleagues published the findings from studies they conducted to examine the impacts of educational technology (O'Dwyer, Russell, & Bebell, 2004; O'Dwyer, Russell, Bebell, & Tucker-Seeley, 2005, 2008). Capitalizing on the hierarchical structure of the data, the authors were able to disentangle the student, teacher, and school correlates of technology use and achievement. For example, the authors found that when teachers perceived pressure from their administration to use technology and had access to a variety of technology-related professional development opportunities, they were more likely to use technology for a variety of purposes. Conversely, when schools or districts enforced restrictive policies around using technology, teachers were less likely to integrate technology into students' learning experiences (O'Dwyer, Russell, & Bebell, 2004). Looking at the relationship between student achievement on a state test and technology use, the authors found weak relationships between school and district technology-related policies and students' scores on the ELA and mathematics assessments (O'Dwyer, Russell, Bebell, & Tucker-Seeley, 2005, 2008). Of course, as discussed previously, the lack of an observed relationship may be due, in this case, to the misalignment and broad nature of the state test compared to the specific skills affected by technology use.

More recently, a large-scale quasi-experimental study of Texas' 1:1 laptop Immersion Pilot program employed a three-level hierarchical model to determine the impacts of 1:1 technology immersion across three cohorts of middle school students on the annual Texas Assessment of Knowledge and Skills (TAKS) assessment (Shapley, Sheehan, Maloney, & Caranikas-Walker, 2010). Using this approach, the authors found that teachers' technology implementation practices were unrelated to students' test scores, whereas students' use of technology outside of school for homework was a positive predictor. In sum, researchers and evaluators must pay close attention to the context within which a technology program is implemented; statistical models that account for the inherent nesting of educational data and include contextual measures and indicators will provide a more

nuanced and realistic representation of how technology use is related to important educational outcomes.

Discussion/Conclusions

This paper explores some of the common methodological limitations that can pose significant challenges in the field of educational technology research. Individually, each of these concerns and limitations could undermine a study or investigation. Collectively, these limitations can severely limit the extent to which research and evaluation efforts can inform the development and refinement of educational technology programs. The overall lack of methodological precision and validity is of particular concern, given the considerable federal, state, and local investments in school-based technologies as well as the current emphasis on quantitative student outcomes. Many of these limitations contribute to the shortage of high-quality empirical research studies addressing the impacts of technology in schools. Currently, decision makers contemplating the merits of educational technology are often forced to make decisions about the expenditure of millions of dollars with only weak and limited evidence on the effects of such expenditures on instructional practices and student learning.

With the rising interest in expanding educational technology access, particularly 1:1 laptop initiatives, the psychometric and methodological weaknesses inherent in the current generation of research results in studies that (a) fail to capture the nuanced ways laptops are being used in schools and (b) fail to align learning outcome measures with the measures of student learning. Beyond documenting that use of technology increases when laptops are provided at a 1:1 ratio, the current research tools used to study such programs often provide inadequate information about the extent to which technology is used across the curriculum and how these uses may affect student learning.

Although this paper outlines a number of common methodological weaknesses in educational technology research, the current lack of high-quality research is undoubtedly a reflection of the general lack of support provided for researching and evaluating technology in schools. Producing high-quality research is an expensive and time-consuming undertaking that is often beyond the resources of most schools and individual school districts. At the state and federal level, vast amounts of funds are expended annually on educational technology and related professional development, yet few, if any, funds are earmarked to research the effects of these massive investments. For example, the State of Maine originally used a \$37.2 million dollar budget surplus to provide all seventh and eighth grade students and teachers with laptop computers. Despite the fact that Maine was the first state to ever implement such an innovative and far-reaching program, approximately \$200,000—or about one half of one percent (0.005%) of the overall budget—

was allocated for research and evaluation. A surprising number of educational technology investments of similar stature have had even fewer funds devoted to their study.

Recognizing that collecting research in educational settings will always involve compromises and limitations imparted by scarce resources, we suggest that extensive opportunities currently exist to improve data collection and analysis within the structure of existing research designs. Just as technology has transformed the efficiency of commerce and communication, we feel that technology can provide many opportunities to advance the art and science of educational research and measurement. Given that educational technology research typically occurs in educational settings with enhanced technology access and capacity, there is a conspicuously untapped opportunity to employ technology-based tools to enhance research conducted in these high-tech settings. In other words, the educational technology research community is uniquely situated to pioneer technology-enhanced research. However, given the budget limitations and real-world constraints associated with any educational technology research or evaluation study, it is not surprising to witness that so few have capitalized on technology-rich settings. For example, although Web-based surveys have become commonplace over the past decade, few represent anything more than a computer-based representation of a traditional paper-and-pencil survey.

In our own work, we have devised new solutions to overcome the obstacles encountered while conducting research in schools by capitalizing on those technologies increasingly available in schools. In this article, we have specifically shared some of the techniques and approaches that we have developed over the course of numerous studies in a wide variety of educational settings. For example, we have found the visual analog scale to be an improvement over our past efforts to quantify the frequency of technology use via survey. Similarly, we have shared other examples of our struggles and successes in measuring the impact of educational technology practices on student achievement. The examples from the literature and our own examples both serve to underscore how quickly things can change when examining technology in education. For example, the ways that teachers use technology to support their teaching has evolved rapidly, as has student computer access in school and at home. In the coming decades, educators will undoubtedly continue to explore new ways digital-age technologies may benefit teaching and learning, potentially even faster than we have previously witnessed, as the relative costs of hardware continue to decrease while features and applications increase. Similarly, the field of assessment continues to evolve as schools and states explore computer-based testing as a more cost-effective alternative to both standardized and teacher-constructed tests. As schools, educational technology, and assessment all continue to evolve in the

future, new opportunities will exist for researchers and evaluators to provide improved services and reflective results to educators and policy makers.

In closing, it is our hope that the issues this article raises and the specific examples it includes spur critical reflection on some of the details important to data collection and educational technology research. In addition, we hope our own examples reported here also serve to encourage others to proactively develop and share what will be the next generation of research tools. Indeed, as technology resources continue to expand and as digital data collection grows increasingly mainstream, we look forward to welcoming a host of new applications of technology resources for improving educational research and measurement.

Acknowledgments

Some of the research summarized in this paper was supported and conducted under the Field Initiated Study Grant Program, PR/Award Number R305T010065, as administered by the Office of Educational Research and Improvement, U.S. Department of Education. The findings and opinions expressed in this report do not reflect the positions or policies of the Office of Educational Research and Improvement or the U.S. Department of Education.

Author Notes

Damian Bebell is an assistant research professor at Boston College's Lynch School of Education and a research associate at the Technology and Assessment Study Collaborative. He is currently directing multiple research studies investigating the effects of 1:1 technology programs on teaching and learning, including collaborative research with the Boston Public Schools and the Newton Public Schools. His research interests include the development and refinement of methodological tools to document the impacts of educational technology on learning, education reform, testing, and 1:1 computing. Correspondence regarding this article should be addressed to Damian Bebell, 332 Campion Hall, Boston College, Chestnut Hill, MA 02467. E-mail: bebell@bc.edu

Laura M. O'Dwyer is an assistant professor in the Lynch School of Education at Boston College and has contributed to numerous studies that examined issues such as the relationship between tracking practices and mathematics achievement, the impact of a technology-infused professional development program on student and teacher outcomes, the effects of a capacity-building online professional development program on teacher practice, and the relationship between the organizational characteristics of schools and teachers' use of technology as a teaching and learning tool. Correspondence regarding this article should be addressed to Laura O'Dwyer, 332 Campion Hall, Boston College, Chestnut Hill, MA 02467. E-mail: odwyerl@mail.bc.edu

Michael Russell is an associate professor in Boston College's Lynch School of Education, a senior research associate for the Center for the Study of Testing Evaluation and Educational Policy, and the director of the Technology and Assessment Study Collaborative. He directs several projects, including the Diagnostic Algebra Assessment Project, the e-Learning for Educators Research and Evaluation Study, the On-Line Professional Education Research Study, and a series of computer-based testing accommodation and validity studies. His research interests lie at the intersection of

technology, learning, and assessment and include applications of technology to testing and impacts of technology on students and their learning. Correspondence regarding this article should be addressed to Michael Russell, 332 Champion Hall, Boston College, Chestnut Hill, MA 02467. E-mail: russelmh@bc.edu

Tom Hoffmann is a research associate at Boston College interested in interface design with a focus on Universal Design principles and usability. He oversees the interface design and production of iNTASC Research projects involving laptop and Internet data collection. Correspondence regarding this article should be addressed to Tom Hoffmann, 332 Champion Hall, Boston College, Chestnut Hill, MA 02467. E-mail: hoffmath@bc.edu

References

- Angrist, J., & Lavy, V. (2002). New evidence on classroom computers and pupil learning. *The Economic Journal*, 112, 735–765.
- Baker, E. L., & Herman, J. L. (2000). *New models of technology sensitive evaluation: Giving up old program evaluation ideas*. SRI International: Menlo Park, CA. Retrieved January 10, 2003, from <http://www.sri.com/policy/designkt/found.html>
- Bebell, D., & Kay, R. (2009). *Berkshire Wireless Learning Initiative: Final evaluation report*. Boston, MA: Technology and Assessment Study Collaborative, Boston College. Retrieved June 15, 2009, from http://www.bc.edu/research/intasc/researchprojects/bwli/pdf/BWLI_Year3Report.pdf
- Bebell, D., & Kay, R. (2010). One to one computing: A summary of the quantitative results from the Berkshire Wireless Learning Initiative. *Journal of Technology, Learning, and Assessment*, 9(2), 1–60. Retrieved December 28, 2009, from <http://www.jtla.org>
- Bebell, D., O'Dwyer, L., Russell, M., & Hoffmann, T. (2007). *Methodological challenges (and solutions) in evaluating educational technology initiatives*. Paper presented at the Annual Meeting of American Educational Research Association, Chicago, IL.
- Bebell, D., & Russell, M. (2006). *Revised evaluation plan for Berkshire Wireless Learning Initiative*. Chestnut Hill, MA: Boston College, Technology and Assessment Study Collaborative.
- Bebell, D., Russell, M., & O'Dwyer, L. M. (2004). Measuring teachers' technology uses: Why multiple measures are more revealing. *Journal of Research on Technology in Education*, 37(1), 45–63.
- Becker, H. (1994). *Analysis and trends of school use of new information technologies*. Washington, DC: Office of Technology Assessment.
- Becker, H. (1999). *Internet use by teachers: Conditions of professional use and teacher-directed student use*. Irvine, CA: Center for Research on Information Technology and Organizations.
- Burstein, L. (1980). The analysis of multi-level data in educational research and evaluation. In D. C. Berliner (Ed.), *Review of research in education* (Vol.8, pp. 158–233). Washington, DC: American Educational Research Association.
- Cronbach, L. J. (1976). *Research on classrooms and schools: Formulation of questions, design, and analysis* (Occasional paper). Stanford, CA: Stanford Evaluation Consortium, Stanford University.
- Cuban, L. (2006). The laptop revolution has no clothes. *Education Week*, 26(8). Retrieved on October 26, 2006, from <http://www.edweek.org/tb/2006/10/17/1040.html>
- Fedorchak, G. (2008). *Examining the feasibility, effect, and capacity to provide universal access through computer-based testing*. Dover, NH: New Hampshire Department of Education.
- Glass, G. V., & Hopkins, K. D. (1996). *Statistical methods in psychology and education* (3rd ed.). Needham Heights, MA: Allyn & Bacon.
- Goldberg, A., Russell, M., & Cook, A. (2003). The effect of computers on student writing: A meta-analysis of studies from 1992 to 2002. *Journal of Technology, Learning, and Assessment*, 2(1), 1–52.
- Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.

- Haney, W. (1980). Units and levels of analysis in large-scale evaluation. *New Directions for Methodology of Social and Behavioral Sciences*, 6, 1–15.
- Horkay, N., Bennett, R. E., Allen, N., Kaplan, B., & Yan, F. (2006). Does it matter if I take my writing test on computer? An empirical study of mode effects in NAEP. *Journal of Technology, Learning, and Assessment*, 5(2), 1–39. Available at <http://www.jtla.org>
- Kreft, I., & de Leeuw, J. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: SAGE.
- Lerman, J. (1998). *You've got mail: 10 nifty ways teachers can use e-mail to extend kids' learning*. Retrieved January 10, 2003, from <http://www.electronic-school.com/0398f5.html>
- Mathews, J. (1996, October). *Predicting teacher perceived technology use: Needs assessment model for small rural schools*. Paper presented at the Annual Meeting of the National Rural Education Association, San Antonio, TX.
- McNabb, M., Hawkes, M., & Rouk, U. (1999). *Critical issues in evaluating the effectiveness of technology*. Proceedings of the Secretary's Conference on Educational Technology: Evaluating the Effectiveness of Technology. Retrieved January 10, 2003, from <http://www.ed.gov/Technology/TechConf/1999/confsum.html>
- Nunnally, J. C. (1978). *Psychometric theory*. New York, NY: McGraw-Hill Book Company.
- O'Dwyer, L. M., Russell, M., & Bebell, D. J. (2004) Identifying teacher, school, and district characteristics associated with elementary teachers' use of technology: A multilevel perspective. *Education Policy Analysis Archives*, 12(48). Retrieved September 14, 2004, from <http://epaa.asu.edu/epaa/v12n48>
- O'Dwyer, L. M., Russell, M., Bebell, D., & Tucker-Seeley, K. R. (2005). Examining the relationship between home and school computer use and students' English/language arts test scores. *Journal of Technology, Learning, and Assessment*, 3(3), 1–46. Available at <http://www.jtla.org>
- O'Dwyer, L. M., Russell, M., Bebell, D., & Tucker-Seeley, K. (2008) Examining the relationship between students' mathematics test scores and computer use at home and at school. *Journal of Technology, Learning and Assessment*, 6(5), 1–46. Available at <http://www.jtla.org>
- Office of Technology Assessment (OTA). (1988). *Power on! New tools for teaching and learning*. Washington, DC: U.S. Government Printing Office.
- Office of Technology Assessment (OTA). (1989). *Linking and learning: A new course for education*. Washington, DC: U.S. Government Printing Office.
- Office of Technology Assessment (OTA). (1995). *Teachers and technology: Making the connection*, OTA-EHR-616. Washington, DC: U.S. Government Printing Office.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage Publications.
- Ravitz, J., Wong, Y., & Becker, H. (1999). *Teacher and teacher directed student use of computers and software*. Irvine, CA: Center for Research on Information Technology and Organizations.
- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15, 351–357.
- Roblyer, M. D., & Knezek, G. (2003). New millennium research for educational technology: A call for a national research agenda. *Journal of Research on Technology in Education*, 36(1), 60–72.
- Russell, M. (1999). Testing on computers: A follow-up study comparing performance on computer and on paper. *Education Policy Analysis Archives*, 7(20), 1–47.
- Russell, M., & Haney, W. (1997). Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via paper-and-pencil. *Educational Policy Analysis Archives*, 5(3), 1–20.
- Russell, M., O'Dwyer, L., Bebell, D., & Miranda, H. (2003). *Technical report for the USEIT study*. Boston, MA: Technology and Assessment Study Collaborative, Boston College. Retrieved March 11, 2010, from <http://www.bc.edu/research/intasc/library/useitreports.shtml>
- Russell, M. & Plati, T. (2001). Mode of administration effects on MCAS composition performance for grades eight and ten. *Teachers College Record*, [Online]. Retrieved March 11, 2010, from <http://www.tcrecord.org/Content.asp?ContentID=10709>

- Salomon, G, Perkins, D., & Globerson, T. (1991). Partners in cognition: Extending human intelligence with intelligent technologies. *Educational Researcher*, 20, 2–9.
- Shapley, K. S. (2008). *Evaluation of the Texas Technology Immersion Pilot (eTxTIP): Year 2 results*. Paper presented at the 2008 Annual Meeting of the American Educational Research Association, New York.
- Shapley, K. S., Sheehan, D., Maloney, C., & Caranikas-Walker, F. (2010). Evaluating the implementation fidelity of technology immersion and its relationship with student achievement. *Journal of Technology, Learning, and Assessment*, 9(4), 1–69. Retrieved March 11, 2010, from <http://www.jtla.org>
- Silvernail, D. (2008). *Maine's impact study of technology in mathematics (MISTM)*. Paper presented at the 2008 Annual Meeting of the American Educational Research Association, New York. Retrieved March 11, 2010, from <http://www2.umaine.edu/mepri/?q=node/11>
- Strudler, N. (2003). Answering the call: A response to Roblyer and Knezek. *Journal of Research on Technology in Education*, 36(1), 73–77.
- Tucker-Seeley, K., (2008). *The effects of using Likert vs. visual analogue scale response options on the outcomes of a Web-based survey of 4th through 12th grade students: Data from a randomized experiment*. Unpublished Doctoral Dissertation, Boston College.
- U.S. Census Bureau. (2006, August 16). *Back to school 2006–2007: Facts for features*. U. S. Census Bureau's Public Information Office. Retrieved March 11, 2010, from http://www.census.gov/Press-Release/www/releases/archives/facts_for_features_special_editions/007108.html
- Wainer, H. (1990). *Computerized adaptive testing: A primer*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Waxman, H. C., Lin, M., & Michko, G. M. (2003). *A meta-analysis of the effectiveness of teaching and learning with technology on student outcomes*. Naperville, IL: Learning Point Associates. Retrieved April 22, 2004, from <http://www.ncrel.org/tech/effects2/>
- Weston, M. E., & Bain, A. (2010). The end of techno-critique: The naked truth about 1:1 laptop initiatives and educational change. *Journal of Technology, Learning, and Assessment*, 9(6), 1–26. Retrieved March 11, 2010, from <http://www.jtla.org>
- Zucker, A., & Hug, S. (2008). Teaching and learning physics in a 1:1 laptop school. *Journal of Science Education Technology*, 17(6), 586–594.