Evaluating Educational Technology Interventions: How do we know its working.

Daniel Light
Center for Children and Technology,
Education Development Center, Inc.

Paper presented at Quest in Bangalore, India. August 2008
Introduction

This talk describes some of the roles evaluation research is playing in advancing the effective use of educational technologies in the US. As we look towards a future of sharing our experience with colleagues around the world, this talk is an opportunity to reflect on the rich history of the Center for Children and Technology (CCT) and to think about what we have learned about how to conduct effective research, and to consider how we might improve what we do and how we work. My comments in this paper build on our collective experiences as researchers during twenty-one years of investigating how technology can best be integrated into high-quality educational environments. Our discussion emphasizes the importance of locally valid and locally useful research designs and attempts to define our approach to conducting evaluations.

The challenge of combining validity and utility is increasingly at the center of our work at CCT. Specifically, we are seeking to conduct research that will help both the research community and educators to understand how complex organizations, like schools, school districts, state and national educational authorities, finance and implement educational technologies, and how those practices might best be improved. In this paper we argue that effective evaluation must produce both research-based knowledge of what technological applications can work best in various educational environments, and practice-based knowledge of how the technology integration process can best be designed to meet locally defined learning goals in schools.

The first section of this paper is a brief review of the recent history of U.S. research related to educational technologies and some of the lessons we have learned from this work. This review points to some of the promising future directions for educational research. In the second section we specifically discuss a role for evaluation in meeting the challenges of helping educators successfully integrate meaningful uses of technology. The third section discusses an evaluation model that stresses collaborative work between research groups, like CCT, and local educators.

Our strong concern with conducting research that is not only rigorous and valid but also useful to practitioners grows out of our collaborative experiences with educators working in many different settings. The Center for Children and Technology has been asking questions about how technology can best support teaching and learning in K-12 schools and other educational contexts for over twenty years. Our work at CCT brings me into contact
with many different types of institutions: school districts, museums, individual teachers, college faculty members, after-school programs, and many others. These relationships take many different forms, but they always require us to value the needs and priorities of those individuals and institutions that are working with us. Working closely with classroom educators, administrators, policymakers, and curriculum and tool developers has pushed us, as researchers, to reflect on and question our theoretical and methodological groundings, and to be both explicit and modest in stating the frameworks and assumptions that guide us in our work. This work and the work of our many colleagues has led us to our current perspective on what is important about infusing technology into K-12 education. We have learned that when student learning does improve in schools that integrate technology, those gains are not caused solely by the presence of technology or by isolated technology-learner interactions. Rather, such changes grounded in learning environments that prioritize and focus a district’s or school’s core educational objectives (Hawkins, Spielvogel, & Panush, 1997).

At the core of our research agenda is a belief that technology can enhance the communicative, expressive, analytic, and logistical capabilities of the teaching and learning environment by supporting types of communication, analysis and expression by students and teachers that are important in two ways. First, the power of technologies offer more flexibility in undertaking certain activities (like writing, editing or graphing) than would otherwise be possible. For example, advanced telecommunications support dynamic and relevant communication with people outside of the classroom; graphic and image technologies allow students to engage with politically ambiguous or aesthetically challenging visual imagery; and word processing makes revision and reworking of original student work easier. Second, technologies can support the extension of learning experiences in ways that would simply be impossible without technological tools — such as visualizing complex scientific data, accessing primary historical source materials, and representing one’s work to multiple audiences. The increasing democratization of access to technology can also make these learning activities available to all students.
I. Lessons Learned from Research\textsuperscript{1}

Researchers, developers, and local educators have been seeking to define the best roles and functions for electronic technologies in educational settings since computers first began appearing in schools, in the mid-1960s (Cuban, 1986). Early studies emphasized the distribution and emerging uses of the then-new tools in schools, as well as learning outcomes of individual students working directly with machines (Papert, 1980). These studies established a body of evidence suggesting that technology could have a positive impact on several dimensions of students’ educational experiences, and researchers began to identify some of the important mediating factors affecting student computer use. At the same time, other studies demonstrated that the nature of the impact of the technology on students was greatly influenced by the specific student population being studied, the design of the software, the teacher’s practices, student grouping, and the nature of students’ access to the technology (Software Publishers’ Association, 1996). This is a key point for educators that we have known for a long time, but seldom really take into account – the success of any technology project depends on the contextual factors and the alignment between context, technology and goals. A number of comprehensive reviews and syntheses of the research conducted during this period are available (Kulik & Kulik, 1991; Software Publishers’ Association, 1997; U.S. Department of Education, 1996).

By the mid-1980s, the situation was changing rapidly. The combination of computation, connectivity, visual and multimedia capacities, miniaturization, and speed has radically changed the potential for technologies in schooling; these developments made possible the production of powerful, linked technologies that could substantially help address some of the as-yet-intractable problems of education (Glennan, 1998; Hawkins, 1996; Koschmann, 1996; Pea, Tinker, Linn, Means, Bransford, Roschelle, Hsi, Brophy, & Songer, 1999). But, because early studies looked so specifically at particular technologies and their impact, they contributed little to the larger, more challenging project of learning about the generalizable roles that technologies can play in addressing the key challenges of teaching and learning, as well as learning about optimal designs for such technologies. In addition, people began to understand that technology’s effects on teaching and learning could be fully understood only in the context of multiple interacting factors in the complex life of schools

\textsuperscript{1} For a more detailed discussion see McMillan Culp et al, (1999).

**Changes in the questions being asked.**

Implicit in the initial strands of research was an assumption that schooling is a “black box.” Research attempting to answer the question, *Does technology improve student learning?*, had to eliminate from consideration everything other than the computer itself and evidence of student learning (which in this type of study was usually standardized test scores – see Kulik & Kulik, 1991). Teacher practices, student experiences, pedagogical contexts, and even what was actually being done with the computers—all these factors were typically excluded from analysis. This was done so that the researcher could make powerful, definitive statements about effects—statements unqualified by the complicated details of actual schooling.

The studies conducted in this way told educators clearly that specific kinds of technology applications—most often integrated learning systems—could improve students’ scores on tests of discrete information and skills, such as spelling, basic mathematics, geographic place-names, and so on. But these studies were not able to tell educators much about addressing the larger challenge of using technology to help students develop capacities to think creatively and critically, and to learn to use their minds well and engage deeply in and across the disciplines, inside school and out.

Past research has made it clear that technologies by themselves have little scalable or sustained impact on learning in schools. To be effective, innovative and robust technological resources must be used to support systematic changes in educational environments that take into account simultaneous changes in administrative procedures, curricula, time and space constraints, school-community relationships, and a range of other logistical and social factors (Chang, Honey, Light, Moeller, & Ross, 1998; Fisher, Dwyer, & Yocam, 1996; Hawkins, Spielvogel, & Panush, 1996; Means, 1994; Sabelli & Dede, 2001; Sandholtz, Ringstaff, & Dwyer, 1997).

In light of this, researchers are increasingly asking questions about 1) how technology is integrated into educational settings; 2) how new electronic resources are interpreted and adapted by their users; 3) how best to match technological capacities with students’ learning needs; and 4) how technological change can interact with and support changes in other aspects of the educational process, such as assessment, administration, communication, and curriculum development.
Changes in methods and measures

Answering such questions requires examining a range of interconnected resources—including technologies, teachers, and social services—that cannot be isolated for study in the way a single software program can be isolated. Further, the kinds of outcomes associated with changing and improving the circumstances of teaching and learning are much more holistic than those measured by standardized tests of specific content areas, and they require more sophisticated strategies of the researcher attempting to capture and analyze them. To explore how best to use technology in the service of these goals requires looking at technology use in context and gaining an understanding of how technology use is mediated by factors such as the organization of the classroom, the pedagogical methods of the teacher, and the socio-cultural setting of the school.

II. What Evaluation Should Do:


Our experience tells us that continued research in this field needs to focus on improving the circumstances of learning, and on determining how technology can help make that happen. This requires viewing technology not as a solution in isolation, but as a key component in enabling schools to address core educational challenges. A consensus has emerged in the U.S. (Dede, 1998; Means, 1994; President’s Committee of Advisors on Science and Technology, Panel on Educational Technology, 1997; Sabelli & Dede, 2001) that the larger issue to be addressed across a wide range of collaborative research projects is gaining an understanding of the qualities of successful technological innovations as they begin to have an impact within local, district, regional, and national contexts.

Implicit in the kind of contextualized evaluation we are proposing is a rejection of past research models that treated schooling (at least for the purposes of study) as a “black box.” These earlier “black box” studies lack local validity, which is an inevitable result of the emphasis put on maximizing generalizability within the scope of individual research projects. The term “local validity” means that information is relevant to and easily understood by school administrators, teachers, parents, or students reviewing the research findings. Local educators seeking to learn from research are unlikely to seek out the commonalities between the subjects in a research study and their own situation. Rather, they are likely to believe that their school, or classroom, or curriculum, are very different from those addressed in the study being reviewed, making traditional research findings not obviously useful to them.
Educators need information about how educational technologies fit in with all the constraints and priorities facing a classroom teacher on any given day. These are precisely the aspects of the research environment (i.e., the classroom) that traditional research models exclude from study (Norris, Smolka, Solloway, 1999). What educators are looking for is not a theoretical understanding of educational technologies or a set of generalized principles about what technology can do, but a contextual understanding of the particular conditions of the implementation, and the contextual factors that interacted with the intervention, that lead to a specific outcome. This is the information they need to find in evaluation research, in order to begin adapting a particular technology to their school and context.

Schoenfeld, a former president of the American Educational Research Association addresses the same concern in a broader discussion of educational research in general. He describes the need “to think of research and applications in education as synergistic enterprises rather than as points at opposite ends of a spectrum, or as discrete phases of a ‘research leads to applications’ model” (Schoenfeld, 1999, p. 14). Schoenfeld highlights the value of creating a dialectic between research and practice, and the need for better theoretical understanding of the complex social systems interacting in educational systems and better conceptualization of the objects of study in research (such as curriculum, assessment strategies, and processes of change [Schoenfeld, 1999]).

At CCT, we argue that this need for a new, more dialectic research framework can best be met by linking together the knowledge-building enterprise of research and its application to the challenges of educational practice, through a research model based on the tradition of evaluation. CCT divides evaluation into two categories according to the questions they pose: formative evaluations examine issues of how and why technology projects work and diffuse within a context or environment; and summative evaluations look at issues of what impacts a project has or how much it changes students’ educational experience. The next section of this paper will present some of the qualities we believe are crucial to designing effective evaluations that can meet our two goals of validity and utility.

**Advantages of evaluation**

Building an evaluation into any educational technology project can have strong implications for the long-term success of the intervention. Including this type of evaluation in an implementation project offers three key advantages. First, any technological intervention into a complex system, like a school or education system, is going to encounter
obstacles, and uncover unexpected opportunities. An evaluation can help to identify and understand both of these possibilities and, often, the evaluator can provide guidance. Second, with this evaluation, the clients (people implementing the technological intervention) have a chance to discuss and shape the evaluation design. This ensures that the evaluation, particularly if it reaches summative phases, meets their needs. Third, this model of evaluation, because it can be attuned to the larger educational context of the project, allows for an exploration of the intervention as a catalyst for change within the larger system.

**How CCT designs an evaluation**

**Framing of the evaluation**

There are two central tenets at CCT about what an evaluation can and should do that provide the intellectual framework for our work in this area. First, CCT firmly believes that an evaluation is an opportunity to establish the terms of success for the technology intervention. We frequently work on projects that begin with unrealistic or oversimplified goals associated with the particular technological intervention. The evaluation process allows the project managers to refine their goals as they gain a better understanding of how their particular project is actually unfolding in practice. An on-going evaluation creates a feedback loop of timely information that allows the project implementers to see emerging problems and develop solutions that help ensure the long term success of the project. The interaction creates the dialectic between research and practice that Schoenfeld feels is urgently needed. The second tenet is a strong belief that evaluation must be carried out over time simultaneous to the different phases of a project. The overall success of any complicated project is dependent on the success of each phase along the way from the initial beginnings to intermediate use to mature use. To truly understand the entire process of a technology project, the evaluators must observe and understand each step of the way.

When implemented, these two ideas of good evaluation design are interrelated in a way that augments the impact and utility of the evaluation. The on-going exchange of information and experiences between the implementers and evaluators at each stage of the project creates the opportunity to rethink and improve the project design along the way.

The on-going feedback between evaluation and implementation is beneficial for the evaluators as well. The mutual sharing of knowledge allows the evaluators to adjust the evaluation plan to capture emerging or unexpected developments. It also allows us to
improve our own critical research skills and deepens our knowledge of the entire process under examination.

**Building a shared understanding of the projects goals**

The first step in designing an evaluation is to begin a conversation with the partner organization that is carrying out the technology project about their goals and objectives (both long and short term), their current implementation strategies and their understanding of the context of the project. These conversations help the organization clarify its goals and expectations, especially if the organization is new to technology, but they also help the organization to begin to understand the role the evaluation can play in helping them achieve these goals. Likewise, these conversations are crucial for the evaluator, as they help us to understand the real issues and goals for this particular technology project. From these conversations, the evaluator can begin to identify the best questions to ask in the evaluation and the best indicators of success to measure.

This process of identifying the right questions to ask in an evaluation builds from our past evaluation experiences and our generalized understanding of some of the key issues that need to be investigated in any evaluation of an educational technology project. We use these previous experiences as a filter, or point of reference, as we begin to shape a new evaluation project. At CCT we believe that the most useful evaluation is produced by researchers who are asking questions about:

- How technology is integrated into educational settings;
- How new digital resources are interpreted and adapted by their users;
- How best to match technological capacities with students’ learning needs; and
- How technological change can interact with and support changes in other aspects of the educational process, such as assessment, administration, communication, and curriculum development.

Questions like these direct the evaluator’s attention to the crucial contextual issues that will, in the end, contribute to a successful technology intervention. Without including these issues in the evaluation study, we have learned, we are likely to misperceive the real value of a particular technological intervention. For example, our experience, and that of our colleagues, has taught us that technology tools can enable good teaching but the tools themselves do not teach. Consider the use of drawing and graphics software, which is common in many U.S. schools. Drawing is a valuable creative activity that students enjoy,
but it is not necessarily supporting the curriculum. In a workshop CCT conducted at the Museo de Arte y Diseño Contemporaneo in San Jose Costa Rica, we used a simple curriculum with driving questions and the drawing tool in Word to engage children in investigating their surroundings, identifying problems and designing solutions. One young girl of about 10 drew her room, planning solutions for a storage problem she had. A secondary student used the drawing tool to develop her designs for a machine that would water houseplants. In this example the interesting elements are not the tool, but how students are asked to use it, how and whether the tool's capabilities correspond to the students' needs, and how and whether the tool is integrated into the overall pedagogical goals for the learner. In this case, the drawing tool allowed the children to express and analyze their own ideas – a substantive learning goal was met well through the use of a technological tool, but it was not the tool itself that caused that learning – it was simply a crucial element of a well-designed learning experience. However, an evaluator seeking to make generalized statements about drawing tools and their impact on learning who did not pay attention to the pedagogical and curricular context of their use would not necessarily uncover the difference between these two situations we have described.

Another topic of these early discussions in an evaluation is scheduling the opportunity for the evaluation to provide feedback to the implementing organization. Evaluations that come at the end of a project arrive too late for the information to have an impact on the long-term success of the project. Given the overall cost of most technology projects, it is wasteful not to have on-going evaluation and feedback. An ideal situation is for the evaluation design to include opportunities for feedback at each stage of implementation, and for even more frequent informal feedback. Ongoing feedback can help implementers correct growing problems before it is too late. Evaluation can also help organizations realize which of their goals are misguided or unattainable, or realize the value of other consequences of their projects. As stated earlier, any intervention into complex systems will encounter unforeseen obstacles and unexpected opportunities.

In one of CCT's evaluation projects, Adventures in Supercomputing, financed by the US Department of Energy, it became clear early on that the Department of Energy had two distinct goals: 1) for students to learn advanced programming languages like Pascal; and 2) for students to undertake complex science projects. The initial project design called for students to learn programming in order to create the computer models to explain their
science project. The time demands to learn complex programming conflicted with the time students needed to research their science projects. In schools, on their own the teachers resolved this problem by reducing the amount of Pascal being taught, and allowing students to build their science projects using spreadsheets, graphs or modeling software (Model It!). During the evaluation, the project coordinators from the Department of Energy came to understand the conflict inherent in their two original goals. They came to agree with the classroom teachers that the having students doing complex science projects was more important than having students learn programming.

The level of analysis: the challenge of evaluating educational technology projects

The model of evaluation that CCT has been developing through our work attempts to overcome an essential challenge in analyzing complex technology interventions into complex systems: identifying the appropriate level of analysis. To provide useful information, the evaluation needs to focus on the right level of analysis, which could be the tool and the learner, the classroom, the school, or the larger school system (such as a school district). Historically, in the "black box" studies, the level of analysis was the tool and the student, since the original need was to establish that technologies could be effective learning tools for students. As the education community confronts the new challenges of promoting wider integration of technology to improve the education throughout schools, we need to shift the level of analysis and are increasingly being asked to do so by our funders. Asking if students learn more grammar or more facts with or without technology, for example, ignores the real issues of technology use as well as the best potential of educational technology. Good technology use tends to transform the learning experience and the learning environment, not merely increase or replace the current learning environment.

For example, the introduction of word processing does not directly translate into better grammar. As a matter of fact, it may move students’ learning away from traditional rote grammar exercises. The best uses of word processing are to facilitate the writing process, create more opportunities for students to write, revise and share their work which will improve their communication skills and their fluency with the written language. They may, therefore, spend less time doing rote exercises. It may also create more opportunity and motivation for teachers and students to learn and practice correct grammar. Students may begin writing long complex papers discussing complex ideas that indicate the development of critical thinking skills, but show no change in their punctuation skills on rote tests.
Writing Matters is a curricula using free online tools to engage students in writing in different genre – from poetry to argumentation. What is the appropriate level of analysis - how student use of the online interface? Or how students as a class engage in writing?

Because of all of these possible permutations of constancy and change, a researcher needs to decide whether it is most appropriate to study individual students’ use of the word processor, to study writing practices within a classroom, or to look at the experiences of many classrooms participating in particular technology-rich writing practices. Which level will best help us to capture both valid and useful knowledge about how the word processor can and do change students’ writing?

The challenge of selecting the appropriate level is also heavily informed by political pressures -- by the tension between the desire of the public, government officials, or international funding organizations for simple conclusions about impact and the real complexities of technology and education. The evaluator is caught with in this tension between simple understandings and complex realities, and choosing a level of analysis that will satisfy public demand for generality while remaining locally valid and useful is difficult, but crucial to conducting useful and effective evaluation. In the hypothetical case of word processing mentioned above, the appropriate level of analysis might not be the individual learner, but the classroom environment and the teacher’s pedagogy, as in most cases, the effective integration of technology requires a change in pedagogy.

Key factors that should be included in the evaluation

Earlier in this paper we discussed some of the factors that we seek to include in the guiding questions of our evaluations. Even though most technology projects focus on specific technologies and applications, we have learned through our work with a variety of schools that numerous social factors influence a school’s ability to use technology effectively for student learning. These factors include:

• Leadership and vision at multiple levels of the system,
• School- and district-wide goals and expectations for the use of technology in the classroom context,
• School culture and climate,
• Teachers’ beliefs about students and their potential for learning,
• Ongoing professional development for teachers.
To fully understand technology interventions, the relative presence or absence of these dimensions needs to be taken into consideration when designing an evaluation. An effective evaluation should be able to situate the technological intervention within these dimensions, and to track changes along these dimensions. We have collaborated with a foundation in Argentina to study the success of certain schools in Red TELAR (a government-financed network of Argentine public schools) (Light, Vilela and Manso, 2001). Through our research we discovered how each school had been able to address each one of these dimensions to make up for lack of effective government support. This type of information would be helpful for the redesign of the program.

An evaluation can also identify when crucial elements along one dimension are missing. For example, in research on a one-to-one computer initiative at an urban high school in the US, it became clear from interviews that students, teachers and school directors each had a different understanding of the vision and goals of the project. As a result the students had no idea of what their responsibilities and were, therefore, failing to meet them. After sharing our findings with the directors of the project, they were able to hold a meeting with students and teachers to clarify the goals of the project. The students are now fulfilling their responsibilities.

Looking for outcomes: definition of success

For an evaluation to be useful, it needs to be built around an expected outcome, some prior conception of what that outcome would look like and a set of indicators of success. This stage of the evaluation entails transforming the general goals and objectives of the project into observable and measurable phenomena. The choice of outcomes is closely linked to the level of analysis: the outcome has to correspond to the level of analysis. It is crucial that the definitions of success be realistically based on the context and a serious appraisal of the project design. The evaluation design must consider that schools are full of complex political and social dynamics. For example, it is unrealistic to expect a project to transform teacher practice if the design does not include professional development nor directly engage classroom teachers.

What kind of evidence

The complexity of schools also means that there are many sources of evidence and information about what happens in these fascinating places. The simplest evidence, and
often the least informative, are test scores. Knowing that mean scores on a standardized test for a class of students went up, for example, does not explain very much because the numbers can not explain how that result came to be. This type of study would also not help understand what aspects of the program promoted success and, therefore, what are key aspects to replicate if the program is extended to other schools, districts or states. To illuminate what has actually changed in a classroom, the evaluator needs to move beyond test results. Other sources of evidence include: class observations, interviews, self-reports (journals, on-line discussions), student products, surveys, and electronic records of activities. In the word processing example above, the appropriate evidence might an examination of student papers, not just students’ test scores.

Quantitative and qualitative methods

The diverse mix of research strategies that CCT uses reflects our belief in the need to combine qualitative and quantitative analysis. A high level of consensus has been reached among educational researchers that both qualitative and quantitative data are necessary in order to generate a complete picture of the impact of a complex initiative on a school community, and that those data need to be coordinated and interpreted in relation to one another. Classroom observations, interviews with teachers and administrators, surveys of students, teachers and parents, and student achievement data all need to be looked at together in order to understand the relationships between quantitative changes that may be occurring and the contextual factors that define, drive and make possible those changes.

As an example here we can cite a project using laptop computers to deepen technology integration in an urban high school. From our survey data we know that the use of presentation software in classes has more than doubled in one year, and that half the students now use PowerPoint. But, our observations and interviews explain how and why this has occurred. To understand how this change come about, we observed the librarian helping students prepare short multimedia presentations using PowerPoint about an academic topic of their interest. The students were encouraged to show their finished presentations to their teachers. In this way, the teachers were introduced to the potential of this tool and saw how competent their students really were with the technology. The teachers began to assign presentations for all their students. The librarian’s initiative, then, was crucial to causing this change. The reasons why the teachers so quickly adapted this technology came out in the interviews. Teachers see these presentations as a powerful
improvement on the traditional written report. The teachers perceive that the ease of the technology allows the students to focus on the content of their presentation, the fact that the students present to each other increases their motivation and competition pushing them to draw on more sources and demand better information, and the multimedia nature of the tool allows students to demonstrate a wider range of abilities.

In this talk, I have tried to present the key components of an evaluation design that can offer useful information for program improvement and measuring the impact of technology interventions. My central argument is that context matters! And the key points are to develop a clear understanding of the program design, agree on meaningful evaluation questions, establish reasonable indicators of success, and target the evaluation at the appropriate level of analyses.

**Bibliography**


http://www.edc.org/CCT


http://www.spa.org/publications/catalog/