Assessing the Impact of Technology in Teaching and Learning

A Sourcebook for Evaluators

Editors

Jerome Johnston and Linda Toms Barker

Published By

Institute for Social Research • University of Michigan

Jerome Johnston and Linda Toms Barker
Assessing the Impact of Technology in Teaching and Learning: A Sourcebook for Evaluators

Published By
Institute for Social Research • University of Michigan

With Funding From
Learning Technologies Division
Office of Educational Research and Improvement
U.S. Department of Education

First Printing: April, 2002

The preparation of this book was underwritten by the U. S. Department of Education, Office of Educational Research and Improvement, Learning Technologies Division. The contents do not necessarily represent the positions or policies of the funder and you should not assume endorsement by the Federal Government.

Table of Contents

Foreword

In 1994 Congress passed the Improving America's Schools Act and a subsequent appropriation that included $45 million to enhance the use of educational technology in American schools. In the ensuing five years Congress steadily increased the funding for technology until, in 1999, the appropriation was $765 million. As they increased funding for technology, Congress asked for better evidence that the investment was having a measurable impact on America's students.

Within the Department of Education, OERI was responsible for three programs that accounted for about one-third of the total technology appropriation: Star Schools, Technology Innovation Challenge Grants, and Regional Technology in Education Consortia. By the year 2000 OERI was monitoring 130 grantees. With each new competition for funds in these program areas, OERI increased the emphasis on evaluation and refined the evaluation requirements in ways that would insure that OERI would secure the information required by Congress when the projects were finished. More than just accountability, OERI wanted to inform the nation about how best to use technology in the schools.

OERI recognized that increasing the evaluation requirements on paper would not automatically yield better evaluations. A consensus needed to be built among grantees, evaluators, and OERI program monitors about what good evaluations looked like. Among evaluators there was little agreement about what constitutes effective research designs or how various concepts (e.g., technology integration, changed pedagogy, learner achievement) should be measured. To address this problem, Cheryl Garnette, Director of the Learning Technologies Division of OERI, saw a solution in convening evaluators for an extended retreat where they could learn about the Department's evaluation requirements and share with each other insights about how best to meet the requirements. More than a sharing, she saw the importance of providing opportunities for evaluators to enhance their methodological skills so they could gather the evidence needed by the Division. With these goals in mind, Garnette authorized the first Technology Evaluation Institute in 1999. Held at the University of Michigan, it lasted three days and included 26 sessions on topics ranging from causal mapping to improving the wording of survey measures. Each session was led by an expert in the evaluation community assigned to review the topic, identify best practices, and lead a discussion around the key points.

The first Institute was a success and provided valuable lessons for all who attended, but there was no documentation of the lessons learned. For the second Institute, held the following summer, a book was planned well in advance. Authors were identified well ahead of time who would take the responsibility for reviewing a key topic area (including reviewing measures used in the evaluations of many of OERI's funded projects), present their review at the Institute, and then revise their review to reflect the comments and insights of those who attended their session. This Sourcebook is the result. It is designed to be a resource for individuals who are conducting evaluations of technology projects. It includes concepts, strategies, and ideas to stimulate the reader's mind about how rigorous evaluation activities, applied early in the process of a project's development, can lead to useful results about student outcomes (including achievement), teacher practices and behaviors, and school climate when technology is implemented effectively.

In spite of the wealth of knowledge that the Sourcebook provides and the experience upon which it is based, it does not serve as the single authority on evaluation strategies to assess technology. It does, however, contribute to the growing knowledge base that serves to inform the education community and helps to link what is learned from evaluation with educational practice. The book has broad implications beyond the review of federally funded projects, although much of the information reported herein was gleaned from efforts supported in whole or in part by the Department of Education. It is intended specifically for use by the community of evaluators and educators who are concerned with assessing the role of technology in American education.

Introduction

Linda Toms Barker and Jerome Johnston
Berkeley Policy Associates • ISR, University of Michigan

Since 1989 the U.S. Department of Education has invested close to a billion dollars in experiments to find compelling uses of technology in public education. The rationale has varied from simply preparing students to function in a technology-rich society to improving instruction of traditional school subjects. If the Department's experiments are going to provide lessons for educators, careful evaluation of each experiment is required. This sourcebook is a resource for the community of evaluators involved in evaluating the more than 100 projects funded by the Star Schools and the Technology Innovation Challenge Grants (TICG). The sourcebook provides an overview of measurement issues in seven areas as well as examples of measures used in current projects. Although designed to address the needs of evaluators of Star Schools and Technology Innovation Challenge Grants it will be of value to the broader community of evaluators concerned with assessing the role of technology in American education.

Background

Given that these technology projects represent a substantial financial investment, it is imperative that OERI be able to report about their success. In the earliest years of the projects, most data being reported were limited to the extent of implementation efforts, with little focus on outcomes or the design features that are associated with success. In 1998 OERI began an effort to enhance the quality of evaluative information coming out of Star Schools and Technology Innovation Challenge Grant projects. A Technology Evaluation Institute was conducted in the summer of 1999 to bring evaluators together from around the country in a formal forum for sharing ideas and experiences in evaluating the success of the grant projects. The institute was so successful that OERI decided to implement another one in the summer of 2000.

At the first Technology Evaluation Institute, one of the most well-received sessions was the Instrument Exchange, in which evaluators were encouraged to bring copies of their evaluation tools to share with other evaluators. Participants went home from the institute with large packets of instruments designed to measure anything from students' technology skills to changes in pedagogy. While this exchange activity was extremely popular and many evaluators reported finding valuable ideas from these examples, this collection of instruments lacked specific context, description, and information about how they were being used. So for the following Institute it was decided to replace this exchange activity with more careful reviews of current practice in evaluating various aspects of these educational technology projects, and the curriculum for the Institute evolved into a more in-depth examination of current practices in the evaluation of technology use in schools.

Thus for the second Technology Evaluation Institute, evaluators were asked to submit example instruments ahead of time so that for each of a number of different measurement areas, a presenter could present background information and describe the range of different approaches currently being used to provide some context for the examples. In the Spring of 2000 six evaluators volunteered to write reviews of current practice in a topic area and to obtain examples of measures being used by Star Schools and TICG evaluators. Each of these presenters agreed to author a chapter for this sourcebook that would define the measurement area, describe the range of different measurement approaches, share examples and reflect on the current state of the art in the specific measurement area.

This Sourcebook of evaluation ideas is an attempt to increase sharing among evaluators of Technology Innovation Challenge Grants and Star Schools projects. This sharing began at a national TICG meeting almost five years ago. Because of the innovative nature of these projects evaluators were eager to learn from each other and spent time together at each national meeting discussing evaluation issues. As the number of grantees and evaluators grew it became clear that a more formal sharing mechanism was needed. What began as one or two sessions at a project director's meeting grew into the Technology Evaluation Institutes at which evaluators conducted workshops and presentations for their peers on various aspects of evaluating educational technology projects.

The authors have laid the groundwork and presented a number of different examples. However, what's presented here is by no means comprehensive. Only a relatively small proportion of TICG and Star Schools evaluators came forth and contributed examples for this first effort. It is our hope that as other evaluators read this work, they will be inspired to contribute their own work to the next version, to make it a more complete sourcebook representing the state of the art in evaluating TICG and Star Schools projects. The ultimate goal is to build toward more common assessments so that the educational community can look across projects and learn what works (and what doesn't) for which learners under what conditions.

What's in This Book?

This sourcebook provides example measures for seven different outcome areas. Each chapter offers first an overview of the measurement area, which includes a review of current relevant literature, and provides a framework for evaluation. This is followed by examples of methods and instruments currently being used in the field. The seven areas included in this sourcebook include:

- Learner outcomes: cognitive domain
- Learner outcomes : affective domain
- Learner outcomes in adult education
- Teacher outcomes : changed pedagogy
- Teacher outcomes: improved technology skills
- Technology integration
- Dissemination outcomes

1. Learner Outcomes in the Cognitive Domain - - *Jerrell C. Cassady, Ball State University.*

The outcome for technology use that would be most valued by educators and parents alike is cognitive gains for students. If students are shown to have increased skills, abilities, performance, or thought processes associated with future success, the innovative educational practice will be judged successful. Even in projects where the goal is to improve the technology infrastructure or enhance teachers' skills with technology, the project is usually justified on the grounds that project activities will eventually lead to gains in student achievement. This chapter reviews the merits, as well as the difficulties, of using standardized tests, tailored tests, and authentic assessments in the evaluation of educational projects. It notes that achievement can be assessed two ways: by direct measures of student learning, or by indirect measures using teachers' and parents' estimates of student learning. Based on the evaluation reports reviewed by the author, the majority of OERI technology projects have used indirect measures.

2. Learner Outcomes in the Affective Domain - - *Jerrell C. Cassady, Ball State University*

Technology projects frequently measure the impact of a project on student attitudes, interests, and motivation for learning. This chapter reviews basic principles of attitude measurement in these areas and provides sample measures used in each of these areas by seven different projects.

3. Learner Outcomes in Adult Education - - *Jerome Johnston, University of Michigan*

Technology projects for adult learners are similar in most respects to technology projects in K-12, but there are important differences in many of the goals, and accordingly, in the measures used to assess these goals. Adult projects include cognitive outcomes such as passing the G.E.D., as well as job application and job performance skills. They may also include goals such as increasing literacy skills (reading a book, completing a workplace form) or enhancing motivation to pursue education beyond the intervention itself. This chapter describes some of the unique outcome measures found in adult education projects.

4. Teacher Outcomes: Changed Pedagogy - - *Marcie Bober, San Diego State University*.

Many technology projects have the goal of changing teachers' pedagogical approach, with the ultimate goal of improving learning opportunities for students. This chapter focuses on strategies for measuring the impact of projects on teacher pedagogy. It includes a review of current thinking about teacher change from the professional development literature followed by examples of technology training efforts that have affected classroom practices as reported by teachers and the students in their charge, and as observed by members of the school community.

5. Teacher Outcomes: Improved Technology Skills - - *Talbot Bielfeldt, International Society for Technology in Education*

Increasing teachers' facility with technology is an important outcome in its own right. Indeed, many argue that without it, students are unlikely to get exposed to technology-based learning experiences in the classroom. What are the technology skills a teacher needs? How can they be measured? What is the relationship between a teacher's technology skills and the integration of these skills into learner-focused instruction?

6. Technology Integration *- - Mario Yepes-Baraya, Harcourt Educational Measurement*

The Learning Technologies Division at OERI envisions the full integration of technology into classroom life. But what does this mean and how can we know when it is achieved? This chapter divides the task into four measurement arenas: the types of technology in use, the primary users (teachers, students, others), the teaching/learning tasks involved, and the cognitive demands required.

7. Disseminating the Lessons of Technology Projects - *- Paula Szulc Dominguez, Hezel Associates*

One important commonality uniting all publicly-funded education initiatives is the expectation that the findings and products will be shared with the education community. Ideally, projects would have generous budgets and expansive timeframes for their dissemination activities, but that is seldom the case. Because of the limited resources available, project staff must prioritize dissemination activities and select ones that offer the most benefits with the greatest efficiency. This chapter provides a framework for planning dissemination activities and a strategy for measuring how well the dissemination effort is working.

What's Not in This Book?

The seven measurement areas included here represent the major outcomes addressed by most TICG and Star Schools project, but are by no means a comprehensive collection of types of measures or evaluation issues. There are a number of other measurement areas that are not included here that have emerged through the evaluation institutes as potential topics of interest and might be a welcome addition to a future edition of the Sourcebook such as:

- Implementation assessment
- Professional development
- Parent/family outcomes
- Community support/attitudes
- Access to technology
- Case study methodology

- Systems change
- Sustainability
- National Performance Indicators: Implications for evaluation.
- Research/evaluation designs
- Technology Tools for Evaluators
- Evaluating Web-Based Interventions/ Estimating Reach
- Program and Curriculum Quality Measures

Observations

From pulling together the chapters for this book and reviewing each author's contribution we can offer a few observations about the current status of evaluation in the TICG and Star Schools programs, at least as revealed in the evaluation materials that evaluators were willing to share with the authors.

First, we notice that, as evaluators, we are still struggling to find a common language to use when discussing the effectiveness of educational technology. Terms such as "technology" and "technology integration" mean different things to different people. Even specific applications such as "word processing" range in meaning from a general notion of text-based communication with computers to specific skills such as creating a newsletter, or entering, copying and moving a block of text. Definitional differences—not only across projects, but within a project between evaluators and developers as well—result in evaluations that are perceived to lack focus, rigor, and usefulness. The Sourcebook takes us one step closer to building a common language for educational uses of technology.

We also observe that most of the measures used in evaluations are "home grown. They were developed by the evaluator to tailor the assessment to the narrow goals of a particular project. The result is measures that directly address the efforts of each grant, but with corresponding limitations in the ability to aggregate or generalize across projects. A greater emphasis on testing the reliability and validity of tailored measures would not only enhance the quality of the evaluations for which they were developed, but also increase the likelihood that they would be adopted for use by other evaluators, thereby increasing the capacity to aggregate results across grantees.

We note also that there is a tendency to focus more on short-term outcomes and effects, rather than seeing the interventions as part of a total package designed to change how schools function. Evaluations are more likely to look at whether students have specific knowledge (standardized tests) than whether students have acquired higher-order thinking and problem-solving skills that would allow them to gain and apply knowledge effectively in the future. Evaluations are more likely to look at whether teachers have mastered specific technology skills than to what extent they can affectively apply those skills to enhancing teaching and learning. Evaluations are more likely to look at math and literacy skills than the kinds of skills that technology may be uniquely suited for such as the ability to predict or simulate. To win acceptance, educational technology is often promoted using an efficiency argument: technology will help students master the traditional goals of schooling (e.g., master math or social studies facts and concepts). These are the goals being measured in existing standardized tests. But technology's strength is often seen in new possibilities for learning: e.g., using simulations to build inferential skills. As the projects mature and new cohorts are funded we see gradual movement towards focusing on longer term outcomes. However, grantees are reluctant to be held accountable for long-term benefits that they may not be able to directly affect during their project period. It will take research efforts beyond individual project evaluations to look comparatively across different interventions designed to address common goals, and to look at the role of technology as part of a larger systemic reform effort.

It may also be outside the capacity of individual project evaluations to develop new measures to assess some of the newer types of outcomes educational programs may want to promote. Thus, while there is still an important need to develop better measures, it may be necessary for OERI to fund separate R&D activities to develop the type of robust measures that could assess these outcomes. One of the most useful contributions of the Sourcebook is the literature review and discussion that begins each chapter and puts the chapter's subject matter in the broader academic discipline. The bibliographies at the end of the chapters are also a useful resource. The appendix provides information on how to contact the evaluators whose work is cited by the authors. These evaluators not only submitted their work and agreed to have it reviewed, but are open to sharing their ideas and experiences with other evaluators.

1. Learner Outcomes in the Cognitive Domain

Jerrell C. Cassady
Ball State University

Learners' cognitive growth is of central interest when evaluating innovative educational programs. If students are shown to have increased skills, abilities, performance levels, or thought processes associated with future success, the innovation is judged favorably. Even when students are not the primary target of the innovation (e.g., a project that provides professional development for teachers or improves classroom Internet access), the project is often justified by the contribution it will ultimately make to student learning.

One indication of the perceived importance of improving academic achievement in educational projects is the Government Performance and Results Act (GPRA), which calls for accountability for the impact of federally-funded programs. Student performance on standardized achievement tests is cited as one of the primary measures of program success. Basic-skills tests of academic achievement are also commonly used at the state and local levels as measures of school or teacher performance (Bauer, 2000). Although standardized tests are effective to a degree, additional strategies used in concert with these measures can provide a more cogent report of a program's impact on achievement.

The goal of this chapter is to describe methods and materials that can be effective in making valid and reliable statements regarding the impact of a particular project on learners' cognitive outcomes. In particular, the areas of academic achievement, critical thinking, problem solving, creativity, knowledge transfer, and skill acquisition are addressed. In addition to a brief clarification of each topic, methodological strategies appropriate for the design of learner cognition measures are described. In all areas of inquiry, evaluators need to match the available methods and materials with the goals and conditions of the specific project under review. As projects differ in scope and focus, several strategies are outlined for each cognitive domain. Finally, examples of measures used in a number of Learning Technologies Division projects are provided, along with information regarding their efficacy and the types of data they yield. (Examples of cognitive outcome measures used in adult education are covered in a separate chapter in this Sourcebook.)

General Concepts

Methodologies available for assessment in the cognitive domains vary widely, with each approach providing a benefit to the overall understanding of the learner. Kennedy's (1999) discussion of levels of approximations of student learning and cognitive outcomes provides a framework for structuring these varying approaches. This level-based classification scheme assumes that the closer the evaluative measure is to the actual work product of the student, the more accurate the approximation of learning or cognition will be. In this manner, direct measurement of performance is the first level approximation with significant others' reports of performance occupying the succeeding levels. Furthermore, within this framework, contextually-based ratings of performance are considered to be closer approximations of learning than general ratings provided through surveys or interviews (Kennedy, 1999). In this chapter, first level approximations will be referred to as direct approximation, with indirect approximation being used to describe all estimations provided by significant others.

Another central theme in this assessment area is the use of multiple methods and data sources to estimate cognitive constructs. Since learning and other cognitive processes cannot be observed directly, experts suggest securing multiple estimates of cognitive constructs from complementary measures (Greene, Caracelli, & Graham, 1989; Popham, 1988).

Finally, the best of measures are of little value when implemented in a weak research design. Collecting comparable measures before and after the treatment, and utilizing both a treatment and control group, can be as important for drawing strong inferences as choosing the right cognitive measure.

Direct Approximations of Achievement

Direct approximations of academic achievement are available from standardized measures, tailored tests, and various alternative assessment methods. Standardized tests of academic achievement that incorporate a national group for normative data are the most highly respected due to the high ratings of reliability and perceived validity of the scales. Performance gains on these scales are seldom questioned as indicators of performance, despite repeated discussions on the downfalls of relying on high-stakes tests as measures of meeting accountability standards (Shepard, 2000).

Standardized tests of achievement that are used at the local and state level are generally written for basic-skills content and are primarily used to indicate proficiency at a particular grade or skill level. However, there are standardized measures of achievement that target higher-order thinking in academic domains. Such measures diverge from basic-skills tests in their topics, intent, and usefulness. However, these tests are far less commonly administered to an entire school or district, leaving basic-skills tests as the more readily available measures.

Standardized Tests

Variability in Administration. One methodological concern with using school or district-mandated, group-administered standardized tests is the variability in administration. Teachers have been observed to vary the methods of test administration enough to significantly influence groups' scores on high-stakes tests of achievement (Wodtke, Harper, Schommer, & Brunelli, 1989). Teacher behaviors that influence student scores include providing additional instructions, rewording instructions for clarity, and answering questions that arise during the test period (violating the standardized task). These variations are generally dismissed as measurement error, but support the preference for using individually administered achievement tests by trained psychometricians when feasible (Aiken, 1996).

Scales. Another common difficulty in making accurate estimates of student achievement is the scale used to report results. Percentile scores are often used to compare the performance of two groups (e.g., Snowman, 1995). However, many analyses using percentile scores are inappropriate due to the lack of interval scaling in percentile scales (Russell, 2000a). Similarly, attempting to make use of raw scores from standardized scales can be misleading. In cases where raw scores are derived from two versions of a test that differ across age groups, the raw scores are not comparable due to the different items, objectives, and tasks inherent in the age-appropriate measures. In situations where the same test items are used across a period of time (particularly with school children), the use of raw scores can also be misleading if the raw score change is not viewed in light of expected growth scores over the given time period (Russell, 2000b; Schulz & Nicewander, 1997). Stanine scores are also problematic in evaluations designed to demonstrate growth or differences between two similar groups, because stanine scores have restricted variability and are discontinuous.

Naturally, the use of a research design that establishes two groups who receive differing levels of exposure to the primary independent variable allows for more confidence when percentiles, raw scores, or stanine scores from archival records are the only measures available. In the case of percentile and raw scores, it is advantageous to transform to a scale score that is universal with respect to standard deviation (Aiken, 1996; Ferguson & Takane, 1989). Suggested scales for these transformations, or for initial collection when available, are z-scores (<u>M</u> = 0, <u>sd</u> = 1.0), T-scores (<u>M</u> = 50, <u>sd</u> = 10), standard scale scores (typically, <u>M</u> = 500, <u>sd</u> = 100), or normal curve equivalence scales using the same scales as the deviation IQ (<u>M</u> = 100, <u>sd</u> = 15). A concern often arises that these standard scales are too esoteric for reporting results to stakeholders. In these cases, it is a simple matter to conduct statistical analyses on a psychometrically appropriate scale and subsequently to transform the results into a format that is accessible to the audience (e.g., percentile scores).

Tailored Tests

Tests that are developed to assess the specific goals of an intervention are referred to as tailored tests. Many educational technology projects use a tailored outcome measure because available standardized tests do not match the goals of the project. Either they do not measure valued goals of the project, or they measure too many goals in addition to those that the project is trying to affect. Often, tailored tests are not as psychometrically sound as standardized measures, because less time and effort is invested at the development stage to refine the pool of items. However, tailored tests can provide compelling evidence when the tests meet basic standards of reliability, validity, and scale construction (Osterlind, 1998). When using tailored tests, include in the evaluation report scale- and item-level information that will establish the quality of the measure (e.g., internal reliability, item-level difficulty and discrimination indexes). (See Osterlind, 1998, for a detailed discussion of test and item analyses). In the absence of these data, the validity of reported outcomes may be dismissed by critics (Aiken, 1996).

Authentic Assessments

Academic achievement estimates can be obtained through authentic assessments as an alternative to standardized tests, typically providing a more context-dependent rating of achievement. Authentic assessment methods typically involve direct examination of work quality through project-based learning, observing performance in a structured setting or presentation, or evaluating work products contained in a portfolio. One alternative strategy to measuring academic achievement relies on systematic observations of student responses to educational or learning antecedents. This approach requires more in-depth analyses of individual learners (thereby limiting sample sizes in most cases), but has an advantage in the standard unit of measure for learning, which is an observable and quantifiable event (Daly, Hintze, & Hamler, 2000).

The guidelines for evaluating work products and performances that do not fall into paper-and-pencil types of measures (or computerized variations of the classic testing strategies) involve making use of established rubrics, relying on multiple artifacts as evidence for performance, and establishing an appropriate level of inter-rater reliability (Hopkins & Antes, 1990). When reporting achievement levels from these more qualitative sources of data it is necessary to provide adequate contextual information concerning the type of assignment, rubrics for evaluating the work, total number of materials available for each learner, and method of selecting the work to be evaluated (Fetterman, 1988; Nevo, 1995).

The use of structured observations by trained evaluators in a classroom setting may also be considered a form of direct approximation of student achievement. As with observations of any behavior, measuring achievement or performance is most effective when the observation goals are clear, observers are trained to a criterion of acceptable performance, and the evaluator has a checklist or scoring protocol to work with during the observation (Hopkins & Antes, 1990).

Indirect Approximations of Achievement

Indirect approximations of learner achievement are estimations of performance level provided by significant others or by records maintained on ratings of achievement (e.g., archived records of report cards). A distinction between grades and performances on achievement measures is made here, as there are generally multiple factors affecting the assignment of a grade by a teacher. Grades are often a combination of measured academic performance coupled with judgments about performance potential (giving the student the "benefit of the doubt"), personal potential, or any one of a myriad of social factors. Teacher ratings of student performance can be gathered through interviews, questionnaires, or archival searches of grade reports or portfolio comments. The more specific the estimate of performance the teacher provides, the closer the rating is likely to be to the learner's actual achievement (Kennedy, 1999). Naturally, the use of teacher ratings adds another dimension of measurement error in the evaluation.

Parents can also provide estimates of their child's performance. The selection of parents over teachers should be viewed with caution, but is appropriate under select circumstances. Parents are better judges of young children's ability levels during preschool years, provided the parents are asked to indicate ability or skill in specific domains and activities. Parents can be more reliable than teachers when the teacher has limited experience with students—for example, at the beginning of a school year. The use of both parents' and teachers' ratings of children's ability or performance levels can enhance the accuracy and quality of performance data.

Assessing achievement levels through self-report is suspect, and falls more into the dimension of self-efficacy ratings (which is addressed in the chapter on measuring learner affect). An exception occurs when asking participants to provide information regarding their performance levels on various tasks (e.g., grades, SAT scores). The reliability of these ratings are generally acceptable for adolescents and young adults, with accuracy of reported scores being significantly influenced by the frequency they have seen the scores and the time elapsed since they last saw the score (Cassady, 2001).

Evaluation of Higher-Order Cognitive Processes

Higher-order cognitive processes generally refer to the wide array of human cognitive activity that goes beyond rote memory and repetition. Evaluations of higher-order processes are often driven by an attempt to characterize learners' cognitive representations of knowledge or learning strategies. The importance of examining data other than achievement measures rests in the theoretical significance of the learning-performance differential. The distinction between learning and performance rests on the fact that all achievement tests are imperfect measures of learning and are thus limited in scope and utility. Thus, there is generally a concern that achievement tests miss the benefits gained in projects that target the development of thinking strategies or skills (e.g., see Kirshstein, Lovitts, Lundmark, & Weidler, 1999). The inferential extensions made in analyses of higher-order cognition are indeed large, and require well-established theoretical rationales prior to making determinations regarding program efficacy.

Whereas the domain of higher-order cognitive processing encompasses a large body of research literature and epistemological theory, this section has been limited to the more common outcome variables addressed in educational programs: critical thinking, problem solving, creativity, and knowledge transfer.

Critical Thinking and Problem Solving

The strategies for evaluating problem solving and critical thinking skills are similar in nature and will be discussed together. However, this is not to assert that problem solving skills and the ability to think critically are evidence of the same cognitive processes. Although they are complementary cognitive activities, and the procedures for detecting ability or growth in these two domains of thinking are often similar, critical thinking and problem solving are distinct constructs.

Critical thinking is one of many cognitive constructs that are ill-defined, or more precisely, differentially defined by researchers from various disciplines and research orientations (Halpern, 1997; Paul, 1990). However, there is some level of agreement that critical thinking is composed of at least two primary components: (a) skill in reasoning and causal assessment and (b) a critical spirit (Siegel, 1992). Critical spirit refers to an attitude that supports critical thought—an attitude of skepticism or belief that conscious effort applied toward thought is worthwhile (Ennis, 1992; O'Flahavan & Tierney, 1991; Sears & Parsons, 1991). The focus on critical thinking in the past two decades derives largely from the judgment by researchers and educators that too many educated individuals were incapable of engaging in critical thought, and a belief that critical thinking skills are necessary for success in modern society (Kennedy, Fisher, & Ennis, 1991). When learners are shown to have heightened skills in critical thinking, and are able to engage in those processes independently, there is heightened expectation the learner will enjoy pronounced success in academic settings, and will likely develop lifelong learning practices and skills.

Whereas the identification of critical thinking generally involves the affective component of attitude, the broad category of cognitive processes referred to simply as "problem solving" is limited primarily to the exhibitions of a combination of skills. Several conceptualizations of the subset of skills involved in problem solving are available, but Sternberg's (1996) is accessible and comprehensive. Sternberg identified six steps that are inherent in most problem solving activities, including

- Recognition of the problem
- Definition of the problem
- Generation of strategies for solving the problem
- Accurate representation of relevant information
- Allocation of resources for problem resolution
- Self-monitoring and evaluation of outcomes.

Problem solving models may also incorporate steps that involve the identification of goals in the problem situation, and an intermediate evaluation of the likelihood of success for generated strategies (Crick & Dodge, 1994). These steps are also commonly tied to a well-established knowledge base that can be referenced in each step of problem solving to identify prior successes and failures, logical outcomes, or similarities of the current situation with familiar ones (Voss & Means, 1989).

The trend to promote project-based learning or guided discovery approaches to instruction has heightened the attention to problem solving skills and critical thinking.

The combination of problem solving and critical thinking strongly resembles the goals of inquiry-based learning strategies, where learners work to answer real-world problems and develop more advanced knowledge bases through their attempts to explain unintuitive results and conditions (for a review, see Chinn & Brewer, 1993). Effective evaluation of higher-order cognitive skills fostered in these inquiry-based learning activities can be facilitated through the acquisition of learning artifacts and sample discourse that demonstrate learner performance in context (Singer, Marx, Krajcik, & Chambers, 2000).

The options available for direct approximations of problem solving and critical thinking are similar to those for achievement testing. That is, there are several standardized performance measures available for both group and individual administration (e.g., Ross & Ross, 1976). These performance-based tasks provide the learner with a series of problem types or questions that require higher-order cognitive strategies for accurate completion. The use of tailored measures of problem solving and critical thinking can be a sensible approach to measuring higher-order cognition, but must be firmly grounded in theory. This is usually attained through task analysis for each set of items and determining the processes most likely to underlie a variety of conclusions or solutions that may be derived in response to the task. Again, artifacts from contextually relevant problem-solving situations are useful materials for making determinations of higher-order cognitive processing, provided the activity leading to the artifact was well-structured (Singer et al., 2000). Technology tools are becoming routinely utilized to gather data on higher-order cognition. Particularly effective methods of gaining insight into learners' cognitive processes, learning styles, and problem solving strategies involve the unobtrusive gathering of real-time data as students work through problems or learning materials (Brem & Boyes, 2000; Jakobsdottir & Hooper, 1995; Welsh, Murphy, Duffy, & Goodrum, 1993).

An alternate strategy for direct approximations of critical thinking and problem solving involves the process of dynamic assessment (Tzuriel, 2000). This approach deviates from static assessment practices by allowing the test administrator to interact with the learner during the assessment period. The test administrator may provide directive feedback or utilize mediational strategies that more directly mimic the learning and performance activities common to educational settings (Grigorenko & Sternberg, 1998). The common static assessment procedures do not allow for interaction between the learner and the test administrator primarily in order to provide testing situations that are identical across settings. However, the use of dynamic assessment strategies has been suggested to provide more accurate estimates of ability for learners who are misrepresented with static assessment tools, including young children, learners with language, developmental, or learning deficits, and children from diverse cultural backgrounds (Tzuriel, 2000).

Obtaining indirect approximations of critical thinking and problem solving is less time consuming due to the relative ease of asking significant others to provide estimations of learners' abilities to perform specific activities considered component processes of critical thinking or problem solving. Self-report ratings of higher order cognitive processes are also feasible, but are generally more effective when the learner provides ratings on distinct skills and abilities, rather than attempting to provide general ratings. Self-reported values are not realistic in young children, as they are not likely to possess the metacognitive awareness or self-appraisal skills needed to provide accurate estimations (Paris & Winograd, 1990).

Creativity

There is little agreement among scholars regarding what constitutes creativity. One attempt stresses that creativity is demonstrated in products or responses that are novel, valuable, and heuristic in nature (Amabile, 1996). But there is some agreement that creativity consists of at least four components: (a) environmental factors that promote creativity, (b) personality characteristics that predict creativity, (c) the production of creative works, and (d) the process of creative thought or action (Taylor, 1988). Promoting creativity in itself may not be as universally valued as boosting test scores, but creative thought is a beneficial cognitive tool for learners to develop and exhibit. The assessment of creativity is tricky, and subject to criticism as there is no standard scale of creativity (Brown, 1989), and developmental issues impact the level of creativity being exhibited (for discussion of this point, see Renzulli, 1986).

Research in the area of creativity has produced varying scales, measures, and methodologies that have attempted to capture the essence of the ability and propensity to produce creative works. Unfortunately, it has been demonstrated that the method of measurement significantly impacts the resultant creativity rating (Parsons, Tittler, & Cook, 1984). In an attempt to categorize the variability in creativity assessment, Hocevar and Bachelor (1989) proposed eight categories:

- Tests of divergent thinking
- Attitude and interest inventories (which may be seen as affective constructs)
- Personality inventories

- Biographical inventories
- Ratings by significant others
- Judgment of products
- Eminence (this method evaluates the characteristics and activities of those judged to be eminent in their field and utilizes these characteristics to evaluate the level of creativity exhibited by project participants)
- Self-reports of activities and achievements

This taxonomy demonstrates the variability in creativity research, but not all approaches provide reliable or valid estimates of creativity. Naturally, it is most desirable to use multiple methods and to provide a more complete approximation of creativity, through triangulation and the selection of complementary approaches (Greene, Caracelli, & Graham, 1989). Of the categories proposed by Hocevar & Bachelor (1989), tests of divergent thinking, review of work products, and ratings by significant others are the most promising and pragmatic approaches for program evaluations.

Direct approximations of creativity can involve the examination of life accomplishments and activities, or the use of an established measure of creativity (e.g., tests of divergent thinking). The relationships between these various strategies generally suggest that established scales of creativity are reasonable predictors of lifelong creative productivity (Torrance, 1988). However, the use of established creativity measures alone may be limited by the loss of contextual relevance (Hennessey & Amabile, 1988).

The shift toward problem-based learning has provided a wealth of finished materials in many educational settings. However, to evaluate creativity in a project, it is necessary to first know the nature of the assigned project. Identical work products can have varying levels of exhibited creativity, depending upon the parameters of the original task assignment. The creation of a computer simulation for a recently read story can be a very creative solution for an assigned book report, but may simply be fulfilling the requirements of an assignment when a teacher specifies the medium and content of delivery. One effective way to evaluate creativity through archival work products is to evaluate several pieces of an individual's work to demonstrate a pattern of creativity. Alternatively, creativity can be assessed through consensus of a group of judges who work independently to assess the creative merits of artifacts by drawing from their expertise in the domain (see Amabile, 1996 for description).

The most common methods of indirect approximations of creativity are through ratings provided by significant others or self-reports of creative activity. Caution needs to be taken when using teacher ratings, as teachers often provide invalid judgments of creativity, and devalue creative characteristics in their students. The research has indicated that teachers typically rate creative behaviors as undesirable student characteristics, providing more favorable ratings for characteristics that demonstrate conformity (Westby & Dawson, 1995). To combat skewed judgments by teachers, measures assessing creativity should solicit specific responses regarding operationalized behaviors, activities, and work products, or simply specify a particular domain of creativity (e.g., verbal) for which the teachers have adequate expertise (Dawson, D'Andrea, Affinito, & Westby, 1999).

Self-reports of creativity are indirect approximations due to the unavailability of the actual creative products or processes for evaluation. However, this approach to developing ratings of creativity addresses the creative person, and identifies those individuals who have produced creative works in various domains and genres. This method of gathering creativity estimations is perhaps the most suspect with respect to inaccurate estimations of what is creative, but can provide an added dimension for creativity when combined with a direct approximation or significant others' ratings.

Transfer of Knowledge or Skill

The transfer of knowledge or skill is a process that is of primary importance in educational projects due to the interest in helping students generalize their learning to settings beyond the initial learning context. Without transfer, most educational or training sessions would be deemed a failure. The ability to transfer learned knowledge or strategies to a novel experience has been linked to the ability for the learner to access the appropriate underlying schema (Gick & Holyoak, 1983), which may be promoted by the recognition of a pattern in the novel situation that matches prior experience (Lockhart, 1992). The ease of transfer of both knowledge and skill depends heavily upon the perceived similarity of the two tasks.

Direct measures of transfer involve providing the learner with a novel problem or situation that resembles the original learning activity in underlying structure, but differs in format, content, or style of presentation. It is valuable to add a measure of transfer of skill or knowledge when measuring achievement, problem solving skills, or critical thinking to demonstrate the breadth of learning obtained in the project. The measurement of knowledge transfer is most reasonable with direct approximation approaches, either through observation of behavior in realistic or contrived performance tasks. Indirect measures may include asking teachers to provide estimates of a learner's ability to transfer knowledge or skill to another task or domain. When this method is employed, it is desirable to solicit knowledge transfer estimates in specific conditions or tasks rather than attempting to gather general ratings of learners' knowledge transfer abilities.

General Principles of Cognitive Measurement Design

Attempting to identify cognitive gains is of primary importance in evaluations of educational projects. However, evaluative methods and materials must align with project goals to provide clear evidence of impact on a relevant cognitive domain. The inter-relationships of the cognitive dimensions identified in this chapter are undeniable, and should be taken into consideration before attempting to evaluate learners in these several dimensions. For instance, it may be inefficient to direct evaluation efforts to extensive measures of problem solving and critical thinking in isolation (unless those constructs are the primary goals of the project). Careful construction of assessment tools can often lead to a single measure providing adequate evidence for multiple cognitive dimensions, ranging from academic achievement to creativity.

The primary concern when measuring learners' outcomes in the cognitive domain is the degree of fit between the psychological construct and the measurement tool. The use of the terms “direct” and “indirect approximation” underscored the state of affairs regarding cognitive measurement. No scale or performance tool is able to entirely eliminate the problems produced by the learning-performance distinction. However, this does not suggest that evaluators should be unconcerned with matters of validity and measurement error. Conversely, the degree of measurement error anticipated for each scale and method of assessment should be a primary factor in the determination of an evaluation plan. As a rule, the use of approaches identified as direct approximations are preferred to indirect approximations when evaluating the cognitive domain. However, triangulation of data, and selection of complementary data sources should be employed when resources are available (Greene, Caracelli, & Graham, 1989).

Evaluators need to employ strategies that limit the threats to validity facing program evaluations that are brought about through the use of tailored tests or measures developed for the unique purpose of evaluating a particular project. First, measures should have adequate reliability and validity prior to making conclusions. In addition to using established measures, reliability and validity can be typically increased by increasing the number of items for each construct (Nunally & Bernstein, 1994). A second effort by evaluators is to gather data on a sufficiently large sample of the target population. The sample also needs to be representative of that population, taking into account as many relevant demographic variables as possible.

Another simple strategy that can be employed to preserve the validity of the data collection process in assessing learner cognition is establishing clear language in all measures. Regarding development, clarity and structure of language is essential to be sure that young children comprehend the intention of the measurement tool. To assure that young children are comprehending the questions, assessment tools can employ concrete examples or manipulatives, or make use of dynamic assessment practices. Another dimension of clarity involves structuring items and measures in the simplest fashion possible for the target audience to limit the number of misinterpretations participants are likely to make. Similarly, the level of specificity for each item helps preserve the integrity of the data collected. As discussed previously, the more specific items are to the target construct, the more likely the answers provided by the participant are valid estimations of the particular cognitive dimension.

Finally, the strength of a solid evaluation plan and research design cannot be understated when attempting to provide compelling evidence of cognitive gains. In the absence of a control group, or at lease baseline data, the efforts toward evaluating achievement, problem solving or any other cognitive activity are meaningless. Furthermore, identification of level of involvement in the target project, and confounding variables needs to be addressed carefully. In projects where educational settings have multiple initiatives functioning simultaneously, the attribution of cognitive growth to the target project is invalid without an adequate comparison group. To accomplish these goals, the evaluator will typically find herself proposing design strategies that put project directors in a position of facing the popular "clinician's dilemma," or battling political pressure to provide services to all learners in an equal capacity. This common barrier to effective evaluation design is best avoided when the evaluator is able to provide early guidance on implementation strategies that follow clear research design, while ensuring breadth of services.

Examples of Measuring Learner Outcomes in the Cognitive Domain

As described above, assessment of learners' cognitive outcomes can be divided into two broad categories: direct and indirect approximations. Both categories include a wide array of methods and materials.

Direct Approximations of Achievement

Direct approximations include achievement tests as well as archival materials, portfolios, and authentic assessment materials or activities. These data can be gathered during the course of participants' involvement in the program, through archival searches, or both.

Achievement Tests

One common type of performance measure used to demonstrate basic-skills proficiency in core academic subject areas is a standardized achievement test. Often, the most readily available measure of basic skills proficiency on a reliable instrument is a state-mandated test, with archival scores. The uses of these tests vary based on the goals of a project; several projects have targeted skills in reading and writing proficiency; others have examined growth in science, while still others are concerned with enhancing student performance across the curriculum. Specific tests that have been used to gather basic information regarding the proficiency levels of project participants have included statewide tests of educational proficiency: California Test of Basic Skills, Terra Nova, and Stanford Nine. The manner of using these tests has varied from tracking a group of individuals throughout a project in a longitudinal analysis of within-group or individual growth, to making use of a control group to examine overall differences based on the intervention.

The **Schools for Thought** project made use of both standardized tests of achievement and tailored instruments to evaluate program impact on learner achievement. The state-mandated Tennessee Comprehensive Assessment Program was found to be less sensitive to the short-term effects examined in the Schools for Thought program evaluation, likely due to the difficulty of one year of instruction to demonstrate significant change in standardized norm-referenced tests (SFT, 1997). The difficulty in demonstrating significant change on the standardized measure was partially attributed to the tendency for scores to be largely influenced by previous years' instruction and socioeconomic status (Meyer, 1996; SFT, 1997). Another troublesome task posed by this design is the difficulty in showing adjustments in norm-referenced scales in a short period of time. Relative measures have expected growth estimations over time, and all individuals should show growth. Therefore, after a one-year period, the experimental group would be required to exceed <u>expected</u> growth estimations to demonstrate any significant impact of the project. This is particularly difficult in a compressed time frame.

The use of existing measures of basic skills provides evaluators with several options when reporting program efficacy. For example, in the **A.C.T. Now! (Anderson, IN)** project, the statewide proficiency test was used to track individual and school-level achievement effects (Cassady, Budenz-Anders, Austin, & Pavlechko, 2001). Because the program services were progressively integrated into schools over time, differential growth patterns in math and English proficiency were available for assessing outcomes. Specifically, effect size gains were noted in the two project schools for the first cohort of students progressing from sixth to eighth grade in both academic areas, while no gains were noted in the control school during the same time interval. In addition to school-level analyses, disaggregation of the data permitted tracking groups of individuals who share unique characteristics, including participation levels, ethnicity, gender, race, and socioeconomic status. The disaggregated analyses permitted more fine-grained analyses of program efficacy, and identification of program elements that were beneficial enough to warrant replication. Because the control school eventually received project services, it is possible to examine growth in those areas since the inception of the program in that site. This quasi-experimental design with multiple baselines provides a strong test of project services.

In addition to school-level analyses, disaggregation of the data allows for tracking groups of students at varying levels of project participation. To assess direct impact of one component of the ACT Now! Project, in which learners were invited to lease computers to keep at home for the school year, individual performance data were gathered over the three-year middle-school period. Group comparisons at sixth grade (prior to computer distribution) demonstrated significant advantages for the control group of students, who were generally not eligible to be participants in this portion of the project due to economic means or because their family already had a home computer. These differences were evident in both the English and math subscales (Cassady et al., 2001). However, after three years of project implementation, the experimental group average was raised significantly, and differences between the two groups were no longer meaningful.

Higher-Order Cognitive Performances

Overcoming the difficulties faced by norm-referenced tests was accomplished in the **Schools for Thought** evaluation by using criterion-referenced performance tests. These tests were constructed to measure academic skills emphasized by the business community, and are included in higher-order cognitive performances due to the focus on real-world problem solving in relation to the academic curriculum (Schools for Thought,

[SFT], 1997). The performance assessments measured skill in reading to evaluate advertisements, writing authentic communication, and complex reality-based math problem solving. For the reading and writing performance areas, the project students responded to one of two potential problem tasks in a pre- and post-test design. The comparison students responded to one of the problem tasks at only the post-test, but were equally distributed across the problem options for each performance area. Data gathering for the math problem solving differed due to access to only post-test implementation. The prompts were:

- Writing authentic communication
  - "If you could change something about your school, what would you change? Why? How?"
  - "If you could change something about your world, what would you change? Why? How?"
- Reading to evaluate
  - Full page newspaper ad with several paragraphs promoting a chemical-based flea trap. Students were asked to summarize the material, define words from context, generate questions they would need for further research, and express an opinion about the ad, while providing evidence for that opinion.
  - Full page newspaper ad with several paragraphs promoting an electronic pest repellant. Students were asked to summarize the material, define words from context, generate questions they would need for further research, and express an opinion about the ad, while providing evidence for that opinion.
- Math problem solving sets
  - Plan a trip, solving distance, rate, time and fuel consumption problems.
  - Conducting a market survey, dealing with statistical sampling and extrapolation from a sample to a population.
  - Interpretation and explanation of information sets represented in complex charts and graphs.

Scoring of these higher-order tasks required ratings from experts. For instance, the writing task was rated on the three-point scale on nine variables: structure, coherence, conclusion, grammar, introduction, length, spelling, transitions, and vocabulary. The outcomes were assessed through expert ratings indicating high, medium, or low competence in several dimensions of the student's response. The design of the Schools for Thought evaluation with these materials enabled two levels of analyses for the reading and writing areas. First, it was possible to measure growth over time in the project participants, which was successfully documented for particular reading and writing skills judged by the expert raters. Second, it was possible to compare performance levels between comparison and project participants on novel tasks at the post-test, which also successfully documented benefits of the project (SFT, 1997).

External Observer's Ratings

The ratings of performance or skill provided by individuals conducting learning or behavioral observations provide a unique approach to assessing ability or performance in a particular context. One organized case study instrument providing such ratings was used for the **ACT Now! Project (Sweetwater Union High School District)** (Lynch, Bober, Harrison, Richardson, & Levine, 1998). In addition to the standard environmental analyses of materials and activities, the case study observation tool calls for an estimate of the students' thinking levels. The levels from which the observer is guided to choose include:

- Mechanical tasks
- Facts
- Information
- Analyze and interpret
- Synthesize
- Judge/assess

Indirect Approximations of Achievement

External ratings of cognitive domains include teacher, parent, or administrator ratings of individual or group benefits in any domain of cognitive growth, including basic skills, technology skills acquisition, higher-order thought processes, and creativity. In addition, observations by external raters can provide data on the skill, ability, or performance levels of the learner.

Teacher Ratings

Using teacher ratings of student ability or performance levels is one of the most common methods of assessing levels of cognitive ability and activity in school-aged children. A common practice for gaining these ratings is to include solicitations for ratings of ability or performance as a subset of items in teacher questionnaires that often focus on several other topics.

Teachers in the **Eiffel Project** provided ratings regarding the expected impact of the project on students, with several items addressing cognitive domains. The teachers were directed to "indicate expectations about the extent to which participating in the Eiffel Project is likely to affect your students' school-related attitudes and behaviors" in each of several domains (Denes, 1999). The rating scale had four points: "none," "a little," "somewhat," and "a great deal."

- Increase content knowledge
- Increase higher order thinking skills
- Increase ability to use primary resources
- Increase ability to function in the workplace.

A similar scale was used in the **Teacher Led Technology Challenge** (TLTC) project (Berkeley Planning Associates, 1999a). The teachers were asked to identify a response indicating "how much difference participating in the TLTC project…made for students" in several areas, using the following four levels: "none," "a little," "a fair amount," and "a great deal." In addition, the teachers were guided to make differential estimations for students who were and were not participants in the target project. Specific items targeting cognitive dimensions include:

- Improved writing/language arts skills
- Improved reading skills
- Improved math skills
- Improved problem-solving skills
- Improved task concentration
- Increased creativity
- Increased technical skills
- Improved ability to manage complex projects
- Improved ability to do research
- Improved critical thinking skills.

In addition to these ratings, the teachers were asked to provide descriptions of ways the project provided positive or negative effects for each student The results from the teacher estimates of cognitive impacts indicated teachers in general found students benefited "a fair amount" in increased technical skills, while the teachers found little or no difference in creativity, problem solving, math skill, or reading skill (Berkeley Planning Associates, 1999a).

Another example of teacher ratings of cognitive outcomes used in the TLTC project is available in a technology skills checklist. The checklist was used for children ranging from kindergarten to second grade. Their teachers rated each child on several discreet technology skills (Berkeley Planning Associates, 1998). The checklist differs from the self-efficacy measures used for the older students, as the ratings are ability ratings that are not influenced by the learner's own affect. The checklist asks the teachers: "based on your observation of the student, please circle the number that best describes his or her ability to complete the following:" The scale had five points: (1) do not know, (2) has never done/cannot do, (3) student can do this with some help, (4) student can do this by his/herself, and (5) student can show someone how to do this. There were 19 items ranging from basic computer operations (turn on the computer, use a mouse, print a file) to various productivity tasks (create a slide show, use a video camera with the computer, use a scanner). The teachers also provided basic information regarding the students including whether students used computers in school, and whether they had a computer at home.

Teachers in the **A.C.T. Now! (Anderson, IN)** project were asked to provide their estimates of the impact of the local project on their students (or general attitudes if the teacher was in a control school). The teachers were provided with a list of items, and were instructed to "check all that apply" to their beliefs that the project computers had a significant impact (Cassady, 2000a). The cognitive dimensions addressed in that scale include student work quality, creativity in student work, and student research.

The **Iowa Distance Education Alliance** used a measure that asked for teacher reports on learners' cognitive outcomes as well. The survey instructions stated: "based on your personal observations, indicate how you feel the integration of technology in learning has influenced students in the following areas." The teachers used a 5-point scale (with an option to also indicate "not sure"): (1) much declined, (2) somewhat declined, (3) no effect, (4) somewhat improved, and (5) much improved (Mauschak, 1999). The items targeting the cognitive domain include:

- Competency in technology skills
- Use of technology in a variety of educational settings
- Acquisition of new technology skills
- Overall academic achievement
- Understanding of "basics" in subject(s) you teach
- Breadth of students' understanding of the subject(s) you teach
- Depth of students' understanding of the subject(s) you teach

- Engagement in activities beyond "rote" learning
- Ability to construct knowledge and solve problems
- Ability to communicate results in a variety of ways
- Engagement in inquiry based learning activities
- Real-life applications of subject

This list of items provides information regarding deeper processing of information, skill acquisition, and transfer of skills and knowledge to multiple realistic settings.

Parent or Guardian Ratings

The **Teacher Led Technology Challenge** also solicited responses from parents regarding information on the cognitive skills and abilities of the students. Parents were asked, "how satisfied are you with each of the following," with a four-point rating scale: (0) very dissatisfied, (1) dissatisfied, (2) satisfied, and (3) very satisfied (Berkeley Planning Associates, 1999b). In particular, parents were asked to rate their child's:

- Reading or pre-reading skills
- Writing skills
- Math skills
- Knowledge of science
- Knowledge of social studies
- Computer skills

Parents provided responses to these ratings in a pre-test, post-test design, which enabled calculation of growth in parent satisfaction with student skill levels. Average satisfaction levels, although they started fairly high at baseline, increased for all academic skills. Not surprisingly, computer skills saw the largest increase, but parents also reported increased satisfaction with their child’s reading skills, math skills and writing skills, and slight increases in satisfaction with social studies and science skills. (Toms Barker,L. Weinstock, P. & Gearhart, M.,1999).

The parents or guardians in the **A.C.T. Now! (Anderson, IN)** project were also asked to provide information regarding their perceptions of the impact of the project on their child's thinking and learning. The parents were asked to provide answers to a list of statements regarding the impact of the project ("this is true for you or your child"). This process involved the simple indication of those items that were true. In particular, three items provided information regarding cognitive outcomes:

- Better prepared for the future due to participation

– Does better in school
– No change due to project

The results from this home-based survey indicated 85% of parents believed their children were better prepared for the future as a consequence of participation in the project. Similarly, over 65% of the parents claimed their children were doing better in school as a consequence of participation, while under 25% of parents indicated that there was no significant impact on their child as a result of participation. (Cassady, 2000b)

Discussion

Overall, it appears that the availability of measures for academic achievement is sufficient, and there are adequate options that are high quality for evaluators to choose from. However, the achievement measures used in many evaluation efforts are not selected by the evaluator because they are already incorporated into the educational environment under review.

As for higher-order cognitive processes, there are significantly fewer established measures used in TICG and Star Schools evaluations. The majority of those measures are indirect approximations of cognitive domains created specifically for limited use in specific program evaluations. This reliance on measures created for specific projects makes comparisons of broad cognitive gains across settings difficult. It would be ideal to have more universal use of established measures in several projects, producing measures with high validity and reliability as well as bringing a higher degree of uniformity in research results for projects with related goals.

References

Aiken, L. R. (1996). *Assessment of intellectual functioning* (2nd ed.). New York: Plenum Press.

Amabile, T. M. (1996). *Creativity in context.* Boulder, CO: Westview Press.

Bauer, S. C. (2000). Should achievement tests be used to judge school quality? *Education Policy Analysis Archives, 8*(46), 18 pp. Available online: http://olam.ed.asu.edu/epaa/v8n46.html

Berkeley Planning Associates (1998). *Technology skills checklist: Grades 6-8.* (Teacher Led Technology Challenge.) Oakland, CA: Author.

Berkeley Planning Associates (1999a). *Follow-up teacher questionnaire: Elementary and middle schools.* (Teacher Led Technology Challenge.) Oakland, CA: Author.

Berkeley Planning Associates (1999b). *Parent/guardian follow-up questionnaire: Elementary and middle schools.* (Teacher Led Technology Challenge). Oakland, CA: Author.

Brem, S. K., & Boyes, A. J. (2000). Using critical thinking to conduct effective searches of online resources. *Practical Assessment, Research & Evaluation, 7*(7), 7 pp. Available online: http://ericae.net/pare/getvn.asp?v=7&n=7

Brown, R. T. (1989). Creativity: What are we to measure? In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 3-32). New York: Plenum Press.

Cassady, J. C. (2000a). *The teacher questionnaire.* (A.C.T. Now! Project.) Muncie, IN: Ball State University.

Cassady, J. C. (2000b). *Impact on families: Spring 2000 A.C.T. Now! Preliminary Report.* Muncie, IN: Ball State University.

Cassady, J. C. (2001). Self-reported GPA and SAT: A methodological note. *Practical Assessment, Research & Evaluation, 7*(12). Available online: http://ericae.net/pare/getvn.asp?v=7&n=12 .

Cassady, J. C., Budenz-Anders, J., Austin, T., & Pavlechko, G (2001). *Integration of computer technology and internet access in schools and homes of children from low-income background: Effects on achievement and attitudes.* Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science education. *Review of Educational Research, 63*(1), 1-49.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information processing mechanisms in children's social adjustment. *Psychological bulletin, 115,* 74-101.

Daly, E. J., Hintze, J. M., & Hamler, K. R. (2000). Improving practice by taking steps toward technological improvements in academic intervention in the new millennium. *Psychology in the Schools, 37*(1), 61-72.

Dawson, V. L., D'Andrea, T., Affinito, R., & Westby, E. L. (1999). Predicting creative behavior: A reexamination of the divergence between traditional and teacher-defined concepts of creativity. *Creativity Research Journal, 12*(1), 57-66.

Delaware Education Research and Development Center (1999). *Delaware Challenge: Year four evaluation report.* Newark, DE: University of Delaware.

Denes, V. (1999). *Eiffel Project teacher questionnaire.* New York: Columbia University, Teachers College, NCREST.

Ennis, R. H. (1992). The degree to which critical thinking is subject specific: Clarification and needed research. In S. P. Norris (Ed.), *The generalizability of critical thinking: Multiple perspectives on an educational ideal* (pp. 21-37). New York: Teachers College Press.

Ferguson, G. A., & Takane, Y. (1989). *Statistical analysis in psychology and education* (6th ed.). New York: McGraw-Hill Publishing.

Fetterman, D. M. (Ed.) (1988). *Qualitative approaches to evaluation in education: The silent scientific revolution.* New York: Praeger.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15,* 1-38.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis, 11*(3), 255-274.

Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin, 124*(1), 75-111.

Halpern, D. F. (1997). *Critical thinking across the curriculum: A brief edition of thought and knowledge.* Mahwah, NJ: Lawrence Erlbaum Associates.

Hennessey, B. A., & Amabile, T. M. (1988). The conditions of creativity. In R. J. Sternberg (Ed.), *The nature of creativity: Contemporary psychological perspectives* (pp. 11-42). Cambridge, NY: Cambridge University Press.

Hocevar, D., & Bachelor, P. (1989). A taxonomy and critique of measurements used in the study of creativity. In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 53-76). New York: Plenum Press.

Hopkins, C. D., & Antes, R. L. (1990). *Classroom measurement and evaluation* (3rd ed.). Itasca, IL: F. E. Peacock Publishers.

Jakobsdottir, S., & Hooper, S. (1995). Computer-assisted foreign language learning: Effects of text, context, and gender on listening comprehension and motivation. *Educational Technology Research and Development, 43*(4), 43-59.

Kennedy, M. M. (1999). Approximations to indicators of student outcomes. *Educational Evaluation and Policy Analysis, 21,* 345-363.

Kennedy, M., Fisher, M. B., & Ennis, R. H. (1991). Critical thinking: Literature review and needed research. In L. Idol & B. F. Jones (Eds.), *Educational values and cognitive instruction: Implications for reform* (pp. 11-40). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lockhart, R. S. (1992). The role of conceptual access in the transfer of thinking skills. In S. P. Norris (Ed.), *The generalizability of critical thinking: Multiple perspectives on an educational ideal* (pp. 54-65). New York: Teachers College Press.

Lynch, E. W., Bober, M., Harrison, P. J., Richardson, H., & Levine, S. (1998). *ACT Now! Case study instruments.* San Diego, CA: San Diego State University.

Mauschak, N. J. (1999). *Status and impact of technology showcase on educational technology current and past winners.* (Iowa Star Schools.) Lubbock, TX: Texas Tech University.

Meyer, R. H. (1996). Value-added indicators of school performance. In E. A. Hanushek & D. W. Jorgenson (Eds.), *Improving America's schools: The role of incentives* (197-223). Washington, D. C.: National Academy Press.

Nevo, D. (1995). *School-based evaluation: A dialogue for school improvement.* Oxford, UK: Pergamon.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd Ed.). New York: McGraw-Hill.

O'Flahavan, J. F., & Tierney, R. J. (1991). Reading, writing, and critical thinking. In L. Idol & B. F. Jones (Eds.), *Educational values and cognitive instruction: Implications for reform* (pp. 41-64). Hillsdale, NJ: Lawrence Erlbaum Associates.

Osterlind, S. J. (1998). *Constructing test items: Multiple-choice, constructed-response, performance, and other formats* (2nd ed.). Boston, MA: Kluwer Academic Publishers.

Paris, S. G., & Winograd, P. (1990). How metacognition can promote academic learning and instruction. In B. F. Jones & L. Idol (Eds.), *Dimensions of thinking and cognitive instruction* (pp. 15-52). Hillsdale, NJ: Lawrence Erlbaum Associates.

Parsons, R. J., Tittler, B. I., & Cook, V. J. (1984). A multi-trait multi-method evaluation of creativity and openness. *Psychological Reports, 54*(2), 403-410.

Paul, R. W. (1990). Critical and reflective thinking: A philosophical perspective. In B. F. Jones & L. Idol (Eds.), *Dimensions of thinking and cognitive instruction* (pp. 445-494). Hillsdale, NJ: Lawrence Erlbaum Associates.

Popham, W.J. (1988). *Educational evaluation* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Renzulli, J. (1986). The three-ring conception of giftedness: A developmental model for creative productivity. In R. Sternberg & J. Davidson (Eds.), *Conceptions of giftedness*, (pp. 74-109). Cambridge, MA: Harvard University Press.

Ross, J. D., & Ross, C. M. (1976). *Ross test of higher cognitive processes manual.* Novato, CA: Academic Therapy Publications.

Russell, M. (2000a). Summarizing change in test scores: Shortcomings of three common methods. *Practical Assessment, Research & Evaluation, 7*(5), 6 pp. Available online: http://ericae.net/pare/getvn.asp?v=7&n=5

Russell, M. (2000b). Using expected growth size estimates to summarize test score changes. *Practical Assessment, Research & Evaluation, 7*(6), 8 pp. Available online: http://ericae.net/pare/getvn.asp?v=7&n=6

Schulz, E. M., & Nicewander, W. A. (1997). Grade equivalent and IRT representations of growth. *Journal of Educational Measurement, 34*(4), 315-331.

Sears, A., & Parsons, J. (1991). Towards critical thinking as an ethic. *Theory and Research in Social Education, 19,* 45-68.

SFT (1997). "Schools for Thought:" Assessment analysis for the 1995-1996 school year. Technical Report. Nashville:

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(1), 4-14.

Siegel, H. (1992). The generalizability of critical thinking skills, dispositions, and epistemology. In S. P. Norris (Ed.), *The generalizability of critical thinking: Multiple perspectives on an educational ideal* (pp. 97-108). New York: Teachers College Press.

Singer, J., Marx, R. W., Krajcik, J., & Chambers, J. C. (2000). Constructing extended inquiry projects: Curriculum materials for science education reform. *Educational Psychologist, 35*(3), 165-178.

Snowman, J. (1995). Computer-based education: More hype than help? *Mid-Western Educational Researcher, 8*(1), 32-35.

Sternberg, R. J. (1996). *Successful intelligence.* New York: Simon & Schuster.

Taylor, C. W. (1988). Various approaches to and definitions of creativity. In R. J. Sternberg (Ed.), *The nature of creativity: Contemporary psychological perspectives* (pp. 99-124). Cambridge, NY: Cambridge University Press.

Toms Barker, L., Weinstock, P. & Gearhart, M. (1999). *Teacher led technology challenge: Year three evaluation report for the Berkeley Unified School District.* Oakland, CA: Berkeley Planning Associates.

Torrance, E. P. (1988). The nature of creativity as manifest in its testing. In R. J. Sternberg (Ed.), *The nature of creativity: Contemporary psychological perspectives* (pp. 43-75). Cambridge, NY: Cambridge University Press.

Tzuriel, D. (2000). Dynamic assessment of young children: Educational and intervention perspectives. *Educational Psychology Review*, *12*(4), 385-435.

Voss, J. F., & Means, M. L. (1989). Toward a model of creativity based upon problem solving in the social sciences. In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 399-410). New York: Plenum Press.

Welsh, T. M., Murphy, K. P., Duffy, T. M., & Goodrum, D. A. (1993). Accessing elaborations on core information in a hypermedia environment. *Educational Technology Research and Development, 41*(2), 19-34.

Westby, E. L., & Dawson, V. L. (1995). Creativity: Asset or burden in the classroom? *Creativity Research Journal, 8*(1), 1-10.

Wodtke, K. H., Harper, F., Schommer, M., & Brunelli, P. (1989). How standardized is school testing? An exploratory observational study of standardized group testing in kindergarten. *Educational Evaluation and Policy Analysis, 11*(3) 223-235.

2. Learner Outcomes in the Affective Domain

Jerrell C. Cassady
Ball State University

The notion that learner affect is a viable factor influencing program efficacy has gained credence, and the need to incorporate the affective domain into comprehensive program evaluations has become evident (Horowitz & Brown, 1996). One clear illustration of the acceptance of affective measures as central outcomes in program evaluation models is seen in the interim project review report from the 1995 Technology Innovation Challenge Grant programs (Kirshstein, Lovitts, Lundmark, & Weidler, 1999). In this report, approximately 50% of the student impact results selected for inclusion in the report were results from the affective domain. The report indicated that project directors and evaluators were uneasy with the practice of relying on reports from standardized tests of basic skills, and reported a strong interest in "improving students' critical thinking and problem solving skills, attitudes towards learning and producing lifelong learners through the reform of the teaching and learning environment (p. 16)." Affective domains addressed as strengths in the interim report indicated success among several projects in learners' (a) classroom participation levels, (b) engagement in classwork, (c) interest in school and learning, and (d) achievement motivation.

The task of outlining all areas of learner affect is beyond the scope and intent of this chapter. Rather, the focus here is on defining affective constructs most commonly investigated and related to the efficacy of educational programs. Specifically, this chapter addresses the conceptual role of (a) learners' attitudes and interest toward learning and learning environments; (b) motivational factors, including goal structures; and (c) self-perceptions of ability. Each construct is defined in the most common ways found in the literature, with the realization that operational definitions for the affective domain vary widely. The chapter then turns to methodological issues and principles that should guide the design of affective measures. Finally, the last portion of the chapter contains a sampling of affective measures used in Star Schools and Technology Innovation Challenge Grant evaluations. This section is intended to help evaluators see how affect has been measured in other projects and then to help with the selection of measures for their own project. At a time when there is extensive experimentation to find the right designs for using technology in the classroom, educators would be helped if there were more commonalities in measuring the outcomes across experiments.

Conceptual and Operational Definitions for Affective Constructs

Affective measures typically serve two primary purposes in program evaluation. First, through descriptive analyses, conclusions regarding the impact of the project on the emotional and motivational state of the participants can be formulated. This common use of affective measures provides quality information about projects and is a promising approach for demonstrating a meaningful effect on the learner.

The second function of affective measures is far less common, but can also provide compelling and essential data regarding program efficacy. This involves incorporating the affective construct in causal models or path analyses, examining the power of the affective constructs as mediators. (See Maruyama, 1998 for extended discussion of these methodologies.) The inclusion of affective constructs as mediator (or moderator) variables strengthens program evaluation in two ways. First, including measures of emotion and motivation in a process model will account for a larger share of the error variance, which will subsequently reduce the probability for Type I error. That is, including learner affect in causal models helps evaluators account for more factors that can explain the variability among project participants. Exploring student affect such as motivation or interest helps the evaluator identify characteristics of learners for whom the project activities were particularly effective or ineffective. Second, the explanatory power of the model is strengthened due to the inclusion of a personality variable that will influence the replication of results in other settings. Thus, it is reasonable that true understanding of the impact of an educational program is only apparent when affective variables have been considered (Horowitz & Brown, 1996).

Attitudes

Evaluation designs that incorporate attitudinal variables typically examine one of two hypotheses. The first is that the project under review will have a positive impact on the learner's attitude toward a component of the program, or some other more distal outcome such as schooling. This avenue of investigation often assesses the learners' interests, which are indicated by persistent positive attitudes toward the identified variable (Snow, Corno, & Jackson, 1996). The second type of hypothesis is that a positive attitude, or high level of interest, toward the independent variable leads to higher personal investment, persistence, continued involvement over time, and enhanced performance (Wigfield, Eccles, & Rodriguez, 1998). In these analyses, differentiation between interest and motivation becomes difficult, as both constructs are measured with similar methodologies, and are influenced by similar processes.

Several evaluation designs involve the use of formative assessments of attitude regarding the state of affairs in the educational setting under review. This formative analysis can be conducted prior to program design in an effort to direct the selection of program goals, or after an initial period of program implementation to assess the strengths and weaknesses of plans and products. These formative attitudinal analyses allow the project to be organic and make adjustments based on the initial reactions of the stakeholders. The validity of these formative attitudinal analyses needs to be examined closely due to the unique conditions that are generally present when formative data are collected. During the initial stage of an intervention, the participants may be overly enthusiastic because of the novelty and appear to be highly invested and engaged in the program. Such "honeymoon effects" often wear off. Conversely, participants may strongly resist the changes imposed by a new program or intervention. In both conditions, the attitudes and emotions of the stakeholders are likely to be tempered with experience, thus the formative data gathered in the first phases of program implementation should not be confused with outcome data.

Generally speaking, summative assessments of attitude are used for three functions in program evaluation. First, learners' interests and attitudes are used to demonstrate the level of congruence between the project's goals and the target population's perception of the project. In this role, attitudes are one measure of goal attainment. The second general use is to determine the overall impact of the project on the learner's attitude toward learning, school, or the particular learning tool or environment targeted in the project. Naturally, in these analyses it is necessary to use an experimental or quasi-experimental design (West, Biesanz, & Pitts, 2000). In the absence of a control group or pretest, the attitudes of learners at the end of a project cannot be reasonably attributed to the efforts of the innovative program under review. The third common function of attitudinal assessment is to strengthen the quality of causal models or path analyses as previously described.

Motivation and Goal Structures

Motivation has been conceptualized in many ways over the course of research in education, learning, and psychology. Early Hullian drive theory defined an unseen intrinsic force that increased the probability of a specific behavior or action (Hull, 1951).

Similarly, operant conditioning models generally hold that there is an internal state of unrest that will drive particular behaviors, with a distinction between primary and secondary motivational forces (see Herrnstein, 1977 for discussion). In a different light, humanistic theories of motivation have defined motivation as the consequence of the need to achieve or avoid stimuli in search of personal fulfillment. Cognitive conceptions of motivation have focused on individuals' attributions for performance (Weiner, 1994), expectations for success (Brophy, 1999), and self-regulation of behavior to meet goals despite internal and external barriers (Schutz & Davis, 2000). Regardless of the underlying theoretical model of motivation, the almost universal assumption is that high levels of motivation leads to higher levels of performance on tasks within an individual's range of ability. Recent research has also demonstrated that motivation mediates the impact of school climate on achievement, when examined in conjunction with the learners' perceived competence (Marchant, Paulson, & Rothlisberg, in press).

One approach to examining motivation is to analyze learners' self-determined goals. Goal structure analyses allow for an indirect evaluation of motivational forces, and can provide insight into task motivation and commitment. There is disagreement as to which is the most appropriate terminology for identifying goal structures held by learners. Three primary categorizations have been proposed and used to describe goals. One classification scheme differentiates between performance and mastery goals. In this scheme, individuals are considered to hold performance goals when the underlying motivation is to perform well on a task in order to preserve self-worth, demonstrate ability and prowess that exceeds peers, or meet a performance standard with little effort (Ames, 1992). Mastery goals generally focus on achieving a high level of performance for the sake of mastering a new skill or self-improvement, with the acceptance that effort has a positive influence on reaching the goal (Ames, 1992). Along the common differentiation of motivational forces, performance goals are seen as extrinsic forces, while mastery goals are more intrinsic in their focus. Two other common categorization strategies involve the differentiation of learning and performance goals (Dweck, 1986) and task-involved versus ego-involved goals (Nicholls, 1984). Despite the variation in label, the differentiation of the goal structures in all three conceptualizations are the same. Dweck's learning goals and Nicholls' task-involved goals have similar underlying motivational explanations as mastery goals, while performance and ego-involved goals are similar to Ames' explanation for performance goals.

Self-perceptions of Ability or Skill

The impact of self-perceptions of ability and skill has been a popular research area since the mid-1970's. The literature on self-efficacy, self-concept, and self-esteem provides evidence regarding the relationships of these constructs with other affective domains, as well as learning and achievement. The distinction among these three areas of self-perception is tenuous at best, as there is often crossover in terminology. One manner of differentiating the three constructs is to identify the level of specificity in self-evaluation. The following description of the self-perception constructs follows a hierarchical, multidimensional model of self-evaluation that was initially proposed by Shavelson, Hubner, and Stanton (1976), and has been continually refined to differentiate varying levels of specificity in self-perception (Marsh & Shavelson, 1985).

Self Concept

Self-concept is operationalized here as a self-assessment of ability based on both internal and external frames of reference (Marsh, 1990b). Self-concept ratings have been found to correlate well with academic achievement, provided the self-concept ratings are not too general in scope. The greatest match between achievement and self-concept is found when the self-concept measure targets specific academic content areas, such as math and reading (Marsh, 1990b).

Despite the fact that self-concept research is almost always based on correlational analyses, the literature has been fixated on establishing the directionality of the relationship between academic self-concept and academic achievement. The notion that self-concept maintains a causal influence over performance is popular and intuitively interesting. This relationship is generally explained through the processes of learned helplessness, motivation or persistence, or self-confidence that allow the individual to perform effectively in evaluative situations (Marsh, 1990b; Schunk, 1990). Unfortunately, it is just as plausible to conclude that a sustained pattern of performance determines the individual's self-perception of ability. Not surprisingly, the data are mixed on the issue, with evidence from different research programs supporting causal paths in both directions (Marsh, 1990a; Helmke & Van Aken, 1995; Rawson & Cassady, 1995). An outcome of the inability to identify a consensus in the directionality of causation between self-concept and performance has been the proposition of bi-directionality (Schunk, 1990). Perhaps the best explanation of the bi-directional relationship of achievement and self-concept is as a spiral effect, in which achievement and self-concept interact through a recursive loop that can manifest as either a vicious cycle of negative decline or progressive acceleration of self-worth and achievement.

Self-Esteem

Self-esteem is typically identified as an individual's perception of her overall ability, or her ability in a particular domain of skills (e.g., academics), along with an evaluative component encompassing how she feels about that ability level (Harter, 1983). The value assessment typically encompassed in ratings of self-esteem, and subsequently the increased level of emotion surrounding self-esteem estimates, is the clearest point of departure from the previous description of self-concept.

The popularity of self-esteem as a construct is undeniable, and is often identified by policy makers and educational stakeholders as a primary concern in education (Kahne, 1996). Interest in self-esteem likely arose due to the widespread acknowledgement of relationships with academic performance, depression, suicide, social ability, and locus of control (Harter, 1983). For this reason, the inclusion of learner self-esteem in evaluation designs is a common practice. However, the evaluation of self-esteem or self-concept should be limited to projects that have an explicit interest in advancing children's self-perceptions, or projects that clearly identify a process through which enhanced self-esteem is expected to facilitate learner success. Although there is a body of research indicating that self-esteem is related to achievement, identifying gains in this dimension should not be treated as evidence that participants have significant gains in learning or achievement.

Self-Efficacy

Self-efficacy is typically conceptualized as a more specific rating of ability than self-esteem or self-concept, and is likened to confidence or comfort with a given activity (Ertmer, Evenbeck, Cennamo, & Lehman, 1994). An individual's rating of self-efficacy is primarily a determination of his ability to complete a specified task under listed conditions, or to "exercise control over events that affect their lives" (Bandura, 1989, pg. 1175). Put simply, self-efficacy is the determination of whether or not the tools and talents necessary to complete a given task are in place.

Ratings of self-efficacy are of interest to educational researchers and evaluators, especially in the field of educational technology due to the high variability in learner skill. Research has consistently identified strong relationships between self-efficacy and (a) persistence (Bandura & Schunk, 1981), (b) intrinsic motivation (Bandura, 1989), (c) self-regulated learning (Schunk, 1998; Stone, 2000), (d) academic achievement (Zimmerman, Bandura, & Martinez-Pons, 1992), and (e) skill with various technological tools (Ertmer et al., 1994). Furthermore, a social-cognitive interpretation of self-efficacy holds that in conditions of low self-efficacy, there is a lower chance that the learner will meet social demands and engage in behaviors necessary to meet external goal standards (Bandura, 1989; Schunk, 1999).

As with self-esteem and self-concept, there is a wealth of research pointing to a persistent relationship between self-efficacy and performance. However, it is important to once again clarify that the vast majority of research on self-efficacy and achievement is based on correlational data, limiting the ability to make determinations of a causal relationship between the two constructs.

Evaluation Strategies and Methodology

Attitudes and Motivation

The methods of assessing attitudes, interests, or motivation are difficult to distinguish from one another due to their construct similarities and the strategies of choice. Thus, in an attempt at parsimony, the program evaluation strategies employed in assessing these three constructs are presented together. Other constructs that follow the methods and procedures shared by these affective measures may include volition, anxiety, feelings of alienation, fear of failure, and goal orientations (see Snow et al., 1996 for detailed discussion).

The process of assessing attitudinal constructs such as motivation, interest, and feelings of comfort (hereafter referred to generally as “attitudes”) differs somewhat from approaches used to evaluate standard cognitive constructs (such as achievement, performance, or intelligence). In particular, the abstract nature of attitudes (and all affective constructs) often leads to a strong reliance on inference. Evaluators often need to rely on inference because (a) attitudes do not lend themselves to direct objective measurement; (b) self-reported attitudes may not be truthful; (c) learners’ attitudes, beliefs, and behaviors do not always align; and (d) attitudes are often measured in a single instance, although they are not constant, universal, or stable traits (Graham & Weiner, 1996; Henerson, Morris, & Fitz-Gibbon, 1987; Wigfield, 1997).

The varied approaches to evaluating attitudes can be divided into four core categories: self-report, significant others’ reports, sociometric ratings, and behavioral observations (for extended discussion, see Henerson et al., 1987). Each methodological approach has both advantages and disadvantages, therefore evaluators should maximize the validity of their data through triangulation, by employing as wide a variety of these strategies as possible (Popham, 1988).

Self-report strategies tend to be dominated by interviews (structured or open-ended), questionnaires, or programmatic journal entries, and require the assumption that the participant is capable and willing to respond adequately and accurately. A more involved investigation of learners' attitudes may be gained through participants systematically documenting their attitudes throughout the course of the day. For instance, participants may write down their feelings about a particular event in a readily accessible log or journal, or make an audiotape recording during the event. When evaluators access these logs, they have more specific contextual information related to the event, providing greater insight and validity than common retrospective or hypothetical measures of affect (Csikszentmihalyi & Nakamura, 1989; Wigfield et al., 1998). Self-reported attitudes are particularly prone to be influenced by social desirability, which threatens the validity of the outcomes. To avoid this threat, the intent of the measurement tool may be masked by asking for ratings of "how most people feel" rather than "how you feel" (Popham, 1988). Alternatively, integrating transparent items or measures into a series of unrelated or distracter items can effectively reduce the ability of the learner to detect the intention of the measure (Henerson et al., 1987).

Gathering data from significant others involves all the methodologies offered through self-report practices, with the inclusion of observational checklists. While significant others often provide more objective reports of affective constructs, they lack the level of insight gained through self-report (Berndt & Keefe, 1995; Marsh Barnes, & Hocevar, 1985). Choice of the appropriate significant other is a major concern, and is influenced primarily by the purpose of the investigation. For attitude assessment, parents and teachers are more appropriate for younger children than peers. However, as children enter adolescence, their peers become more reliable and valid sources of information. In fact, in some domains of investigation such as social competence, adolescents' peers may provide more accurate ratings than parents and teachers .

Sociometric methods of assessing attitude and interest rely upon peer ratings or nominations of individuals with certain qualities or characteristics. Sociometric analyses provide data from multiple sources (typically peers), providing higher levels of confidence in ratings than relying upon single significant other ratings offered by peers. However, reliability declines when the raters are young. Developmental level needs to be considered when making use of this methodology (Furman & Buhrmester, 1992).

Behavioral views of motivation were dominant in the middle of the twentieth century, and produced methods of measuring motivation that relied on observational data (Snow et al., 1996). Although researchers inevitably must make inferences regarding the motivations or attitudes underlying observed behaviors, behavioral approaches to measuring motivation are complementary to self-report and significant others' ratings, and the use of multiple methodologies enhances the accuracy of measurement (Shotland & Mark, 1987). Practices for evaluating motivation through the behavioral paradigm involve observing participants' chosen activities, duration of continued involvement with an activity (particularly those that are challenging or lead to repeated failure), task completion rates, volition, and level of effort exerted by an individual to receive a "reward" (Graham & Weiner, 1996).

An additional set of strategies for gathering data in these four areas involves the use of archival data such as diaries, classroom discipline ledgers, correspondence between school and home, and attendance records (Henerson et al., 1987). These records provide contextual and historical information regarding patterns of learners' attitudes and motivation and provide a historical viewpoint to behavior. However, conclusions formulated from data gathered only through archival searches are likely to be skewed, for instance many archival data only report negative behaviors (e.g., referrals, detention, truancy) and do not have reliable information on rates of desired behaviors. Thus, archival data can provide strong evidence for attitude and motivation, but should be combined with data from other sources.

Self-perceptions of Ability

The measurement of individuals' self-perceptions of ability or competence involves attempting to identify their attributions, perceptions, and expectations for performance on a personal level. Although this personal focus is central to the formation of self-perceptions, the internal representation held by the learner is greatly influenced by external factors, including social interactions, adult praise and reprimand, social comparison, and observational learning (Bandura, 1989; Harter, 1983; Schunk, 1990).

The evaluation of self-perceived abilities differs somewhat based on the specificity of the construct in question. For instance, self-efficacy measurement generally involves providing an example of the target task and asking the individual how confident she is that she can successfully complete tasks of that sort (Bandura & Schunk, 1981; Sawyer, Graham, & Harris, 1992). Self-concept and self-esteem measures that are more general in nature often rely upon hypothetical situations or self-statements that are intended to provide insight into the individual's perceived ability or feeling or worth. These methods are more effective at providing an overall trait value of the concept, but reliance on these broad measures have historically proven to lead to erroneous conclusions of the relationship between self-perceptions and achievement, social functioning, and other affective constructs (Marsh, 1990b; Wylie, 1979)

To overcome the limitations presented by general measures of self-perception, content- or situation-specific measures of self-efficacy or self-concept are used, and are typically combined to form hierarchically superordinate constructs. For instance, several items addressing ability to solve various types of math problems would provide a rating of math self-concept, and a different set of items may combine to provide a rating of verbal self-concept. Following the multidimensional, hierarchical models of self-concept or self-esteem, these two specific areas of academic competence may be compiled to provide an overall academic self-concept or self-esteem rating for the individual. The Self-Description Questionnaire (see Marsh, 1990b) follows this template, by providing multiple dimensions of self-concept for investigation. The number of distinct dimensions forming a learner's self-concept is highly dependent upon development. As individuals progress from preschool to adulthood, their ability to differentiate between multiple "selves" or areas of ability becomes increasingly complex (Cassady, Mantzicopoulos, & Johnson, 1996; Entwisle, Alexander, Pallas, & Cadigan, 1987; Harter, 1982; Marsh, 1989; Stipek, Recchia, & McClintic, 1992). Therefore, the measurement of self-perceptions of ability needs to be tailored to the appropriate developmental age group (see Marsh, 1990b and Stipek et al., 1992 for full reviews).

Another developmental concern regarding the assessment of affect is clarity of language. Various approaches have been employed to ensure younger children comprehend the ratings they are providing. For instance, young children will often have trial items in advance of the actual scale to become familiar with the rating system (Bandura & Schunk, 1981; Cassady et al., 1996). Self-ratings tasks have also been put into pictorial format to overcome language barriers (Harter & Pike, 1984). Another approach intended to reduce the abstract nature of ability ratings is to make use of manipulatives (Cassady et al., 1996). The manipulatives are intended to maintain the attention of young children, as well as provide them with a concrete object that elucidates the range of possible ratings (for instance in making a judgment of ability level from 1 to 5, students displayed the number of stars that demonstrated their ability in various academic subjects). Finally, the use of a two-step forced choice approach has been employed to allow for more valid, fine-grained self-judgments of ability or competence (Harter, 1982; Harter & Pike, 1984). The two-step method works by having children first identify whether a statement is like or unlike them. After making this decision, the children answer a follow-up question that addresses the extent to which the statement is like or unlike them (sort of like me, very much like me; Harter & Pike, 1984). This approach leads to data falling into a 4-point Likert-type scale, with enhanced confidence in accurate ratings provided by young children who otherwise have difficulty differentiating among the 4 points.

One drawback to reliance on self-report data in young children is the tendency for elevated ratings of competence (Marsh, 1989; Stipek & MacIver, 1989). Up through the first few years of formal schooling, children typically report unrealistically high ability estimates in most domains. These elevated levels of perceived ability are often hypothesized to be due to the high level of teacher and parent praise in the early grade levels (Stipek & MacIver, 1989). Inflated self-ratings of ability may serve a positive role in the academic development of young students, promoting persistence over repeated failures in light of the expectation of future success (Bandura & Schunk, 1981). However, this poses a significant problem regarding the validity of assessments of young children's self-perceived competencies. To minimize the threat to validity posed by this inflation tendency, it is often desirable to incorporate significant others' ratings for validation.

The use of significant others' ratings often falls outside the explicit focus of learner affect, if the provided ratings target actual ability. However, the significant others can provide an estimation of what level of self-esteem or self-concept a child will provide (referred to as inferred self-concept; Marsh, et al., 1985; Marsh & Craven, 1991). Ideally, the use of significant others should be used in addition to self-report, rather than replacing the self-report measure, producing a multitrait-multimethod analysis (Marsh et al., 1985), typically referred to as triangulation (Popham, 1988). Selection of significant others is a critical issue in this approach, as parents of preadolescent children are somewhat more accurate in inferring the level of the child's self-concept than teachers, and inferred estimates provided by peers are generally the least accurate (Marsh & Craven, 1991). Difference between parents and teachers in accuracy of inferred self-concept may be due to the longer-term relationship parents hold with the child, or the over-sensitivity of teachers to actual performance levels, which tends to lead to skill ratings rather than inferred self-concept (Marsh & Craven, 1991).

Principles of Affective Measurement Design

The role of learner affect in the outcomes of an educationally innovative project is undeniable, even when influencing affect is not a primary goal. Gathering information on learner affect allows evaluators to progress toward evaluation models that are process-oriented rather than strictly outcome-oriented (Judd, 1987). Affective constructs are commonly identified as mediators, catalysts, or inhibitors to program efficacy (Marchant et al., in press). Thus, even in evaluations of projects where student achievement, skill acquisition, or teacher pedagogy are the primary outcome variables of interest, constructs of learner affect can be identified as necessary, but not sufficient, factors in program success.

Although the use of well-established measures of learner affect is the preferred course of action when attempting to measure the effects of projects under evaluation, it is not always possible to find a measure that directly meets the needs of a specific evaluation. In that case it is necessary to develop evaluation measures using established principles of measurement design. Evaluators should consider the following principles when developing measurement in the affective domain:

- *Domain Sampling:* The measurement of complex constructs requires multiple items to sample the different components of the concept. Avoid using just a few items as evidence of an affective construct: (See Nunnally & Bernstein, 1994 for complete discussion).

- *Social Desirability:* Several strategies exist to overcome the confounding effects of social desirability in the affective domain. Items that are particularly transparent (especially when the sample population involves adolescents or adults) can be masked by integrating them within a set of unrelated items, or placing the transparent items at the end of a longer instrument scale (Henerson, et al., 1987; Popham, 1988).

- *Response Bias:* Variation in the valence of affective items can increase the validity of responses by limiting response set or response biases, where the learner gets into a pattern of responding to the items without reading the individual items carefully.

- *Developmental Appropriateness:* Scales developed for various ages need to use vocabulary and grammatical structure appropriate to the age level. Strategies for young children include: reading the question aloud, using two-step forced choice strategies, use of concrete manipulatives, construct-driven measures that identify age differences in dimensionality (i.e., self-concept), and pictorial cues.

- *Clarity of Language:* Clarity of language is a must for any age levels. It is particularly important for affective evaluation due to the possibility of misinterpretation of the intention of the items. Solutions include pilot-testing the scale with a similar population, integrating multiple phrases that address the same issues into the measurement tool to establish internal consistency in responses, and involving individuals from the locale of the testing to ensure that language tendencies embodied in the scale are represented in the population.

- *Reliability and Validity:* The use of tailored tests in evaluation studies is often unavoidable, but it does not release the evaluator from adhering to established psychometric standards. To establish reliability, the evaluator can at the minimum identify internal scale validity (given enough items), or demonstrate test-retest reliability with a subset of the population. Triangulation generally will provide information useful in making statements of validity.

- *Triangulation:* The use of multiple sources of data, as well as multiple methodologies allows for greatest confidence in identifying the subjective constructs examined in the affective domain.

Evaluation plans that follow these general strategies when assessing the affective outcomes associated with the program under review are likely to avoid several problems that limit the level of compelling data available. By addressing the psychometric and developmental properties of the measurement tools and data collection methodologies in advance, evaluators place themselves in a better position to provide evidence to the stakeholders and the interested professional audience that will help guide future adaptations and applications of the project's core innovations.

The principles above apply to measurement design. If one is to draw conclusions about the impact of an innovation on the affective domain, a rigorous research design is equally important. Pre-post designs and the use of control groups are a necessary complement to good measurement.

Examples from Star Schools and Challenge Grant Evaluations

To elucidate the principles of assessing affect, examples of measures from seven TICG and Star Schools project evaluations are included in this section. These examples may help evaluators select or develop measures for their own project. However, many of the instruments are tailored to unique aspects of projects, and do not translate in current form to all evaluative settings. (Examples of measuring affect in adult education projects can be found in another chapter in the Sourcebook.)

Self-Report Measures of Affect

To evaluate learners' attitudes, interest, and self-efficacy ratings about school and technology, the **Trails Project** made use of a 16-item scale. The instructions for the instrument were “We would like to know your thoughts about schools and computers. Please check the box that best describes what you think about each item.” The students were further directed to respond to the items with respect to “this year.” The scale made use of a four-point scale (Definitely, Usually, Sometimes, Almost Never). For select items, the students were also provided with an option of indicating that the opportunity did not exist for that particular activity—e.g., use a computer to write or publish (Research & Training Associates, 1999; the order of the items is different than in the original instrument).

Self-perceptions

- I do well on most assignments.
- I know how to do research.
- Using computers helps me do better on my schoolwork.
- I can use a computer to locate sources of information.
- I can use a computer to find information that I can use.
- I can use a computer to write or publish.
- I can use a computer to understand data by making graphs and tables.
- I can use a computer to make a presentation.
- I can help other students use computers.

Attitudes & Interests

- I like school.
- I am learning a lot in school.
- I wish I had more chances to use computers in school.
- Schoolwork is more interesting when I work on projects.
- I like working on projects with other students.
- Using computers makes schoolwork more interesting.
- I am learning a lot in this classroom.

The data from use of this instrument indicated positive effects for the Trails Project in the following ways (J. Pfannenstiel, personal communication, August 18, 2000):

- 80% of students reported they *definitely* or *usually* liked project-based learning and working in collaboration with other students.

- Approximately 75% of students reported they *definitely* or *usually* wish they had more chances to use computers in school.

- Approximately 75% of students reported they believe that computers made schoolwork more interesting.

- Student performance gains were noted in learning to do research, locating information with the computer, writing or publishing with computers, and making presentations with the computer

In the **Teacher Led Technology Challenge**, self-efficacy in technology skills was assessed by asking the students to provide ratings of several technology-related tasks, using the following 5-point scale: (1) “I do not know if I have done this,” (2) I have never done this,” (3) “I can do this with some help,” (4) “I can do this by myself,” and (5) I can show someone how to do this.” The instrument was completed by participants in the TechnoKids computer-borrowing program, as well as by non-participants. The students in grades 6 to 8 were asked to rate themselves on all of the skills below; students in grades 3 to 5 rated themselves on a subset of the items (Berkeley Planning Associates, 1998).

- Turn on a computer
- Use a mouse
- Type a story or report
- Edit a story or report on a computer
- Open and close programs
- Print a file
- Save a file
- Rename a file
- Delete a file you no longer need
- Insert and remove a disk
- Copy information onto a disk
- Shut down a computer
- Use E-mail
- Search the Internet to find information
- Create a web page
- Put information into a database
- Look up information in a database

- Create a database
- Create a spreadsheet
- Create graphs and charts
- Use a digital camera
- Use video software to edit video on a computer
- Create graphic designs
- Find out how much RAM the computer has
- Figure out how to use a program I've never seen before
- Draw or paint pictures on a computer
- Record your voice and other sounds
- Create a slide show
- Use a video camera with a computer
- Use a scanner

The Teacher Led Technology Challenge project also measured internalized affect with Harter's Perceived Competence Scale for Children, which provided ratings of children's perceived competencies in the scholastic, social and behavioral domains, as well as a global self-worth rating (Harter, 1982). The results from the self-competence scale demonstrated significant gains by the targeted TechnoKids population, as compared to the comparison group. The results revealed that the TechnoKids held significantly lower views of global self-worth in baseline analyses than the comparison group (for elementary school population). However, after one year of intervention, the TechnoKids ratings of self-worth did not differ from the comparison group's self-worth, which had remained stable across the year (Toms Barker, Weinstock, & Frakes, 2000).

The **Maryland Electronic Learning Community** made use of learner self-report instruments that targeted several dimensions of affect, and attempted to specify the role of the project computers in the affect of the learners through the inclusion of clarifying statements regarding the computers (Mucherah, 2000). The instructions informed the students the survey was "not a test," there were "no right or wrong answers," and that the purpose of the survey was to "learn about what you think about using computers." The learners replied anonymously, with school-level data gathered to compare different program implementations.

Students were asked to respond to a series of items regarding math, science, social studies, and language arts. The students were first asked how often they used computers in those four subject areas (every day, a few times a week, once a week, a couple times a month, or never). Then, the students were asked to respond to several items for each content area on a five-point scale: (1) strongly agree, (2) agree, (3) somewhat agree, (4) disagree, and (5) strongly disagree.

Impact on math attitudes and performance

- I have improved in math as a result of computers at school.
- I have improved in math.
- Since we started using computers in school, I like math a lot.
- I like math.
- [Identical items for science, social studies, and language arts]

Impact on response to school

- Since we started using computers at school, I come to school more often.
- Since we started using computers at school, I like school more.
- Since we started using computers at school, I rarely get in trouble at school.
- I usually come to school most of the time.
- I usually like school a lot.
- I usually don't get in trouble at school.

The structure of this instrument is designed to allow for comparisons between rating levels of liking for a subject area, and the improvement in that attitude since the inclusion of computers in the school setting. Thus, each type of item is posed to the learner as a general statement, and as a contextualized statement with respect to the project (W. Mucherah, personal communication, September 20, 2000).

The **Mississippi Educational Television Project** made use of self-report measures to identify the strengths of the project, and to identify motivation and interest with respect to the unique learning environment posed through their interactive video network (Merker & Associates, 1999a). The learners were asked to use a four-point Likert scale (strongly disagree, disagree, agree, strongly agree) to give their reactions to a number of items that address motivation, interest, and feelings of involvement in the learning process:

- It is easy to pay attention to the TV monitor
- I had a lot of chances to ask questions
- I learned more from the network class than from a regular class

- I feel like I am part of the class even though the students in the other site are far away
- I feel like the teacher is speaking directly to me
- Overall, I liked the video sessions
- The fact that I am "on TV" does not change my class participation
- I don't like to ask questions when I know other students are watching me on the network
- I would like more classes on the network
- With the network classes, I get to see and learn things that I can't usually get in my other classes
- The network classes are more interesting than my regular classes
- I would like more of my teachers to use the network

The responses to the student survey (n = 1,560) were factor analyzed. The first factor (this technology as a preferred teaching method) was strongly associated with overall satisfaction with learning (Merker & Associates, 1999b). Other notable findings include:

- 77% of respondents agreed with the comment that they liked the video sessions overall;
- 70% disagreed with the comment that they were less likely to participate by asking questions while being observed by other viewers; and
- 56% wished more teachers used the network.

The **ACT Now! Project (Sweetwater Union High School)** investigated affect through self report in two ways. The first used focus groups and structured interview protocols. The interview protocol called for the interviewer to identify areas the students particularly liked or disliked and aspects of a recent lesson that were considered to be particularly effective. In addition, a sentence completion task was used to assess attitude toward a recent lesson. The interviewer further addressed the students' attitudes toward the use of technology in lessons (Lynch, Bober, Harrison, Richardson, & Levine,1998). The interviewer asked students to comment on specific lessons that the interviewer observed in previous classroom observations, and to indicate what in particular they learned, liked, or disliked about the lesson. They were also asked to indicate the highlights of the lesson, and what they enjoyed most (Lynch et al., 1998).

The second self-report approach in the ACT Now! project was the use of a student instrument that addressed affect using several different strategies, allowing the evaluator to infer the students' attitudes and perceptions regarding the effectiveness of having the computer involved in their learning experiences (Lynch, 1999). The instrument is broken into two primary areas of inquiry, self-efficacy and attitude. The self-efficacy portions resemble other examples presented in this chapter, asking the children to indicate "how good" they are at using certain software (e.g., web browsers, word processing, presentation). However, this instrument makes use of a four-point scale more commonly used with adults: (1) no experience, (2) beginner, (3) intermediate, and (4) advanced. The meaning of terms such as "beginner" and "intermediate" require precise definition for any respondent. But it would be particularly difficult for students below high school age.

The ACT Now! student technology instrument also addresses attitudes and opinions on the importance of topics associated with the learning environment in the project. One interesting item asks students how often they would like to use computers at school (almost every day, about once a week, about once every two weeks, about once a month, or never), which can be compared to an item regarding how often they actually use the computers in school. Further, the students were asked to indicate "how strongly you feel about" the following items, selecting among three options: very important, somewhat important, and not important at all:

- Using a computer at school whenever I need to
- Logging onto the internet from school
- Learning more about computers and other technologies
- Having other students in my class listen and talk about my ideas
- Helping the teacher decide which activities I do, or which projects I work on

This use of several approaches to assess the affect of the students in the project, including interviews, surveys, and focus groups helps to extend the availability of data that can be used in analyses of affective outcomes.

Another example of using student reports for affective measurement comes from the **Eiffel Project**. The student measures incorporated several ratings of ability, activity choice, and time spent using various materials. The instrument allows inference of motivation and interest in various computer-related activities by asking for estimates of time spent on various activity choices. For instance, the students were asked to report on the frequency of visiting the computer lab on their own time (less than once a month, once a month, 2-3 times a month, once a week, 2-3 times a week, every day). They could also list up to four activities they participate in on their own time. Students provided estimates of the amount of time they spent on various technology-related activities, both at school and at home, using a 5-point scale—none, less than 1 hour, between 1-2 hours, 2-3 hours, more than 3 hours (Denes, 1999a). The activities included:

- Playing games
- Playing educational games
- Visiting chat rooms
- Typing homework assignments
- Surfing the internet
- Emailing
- Researching subjects of interest on the internet
- Researching subjects of interest on a CD-ROM
- Looking for information for school projects on the internet
- Looking for information for school projects on a CD-ROM
- Completing projects and portfolios for school

In addition to these items, the student survey for the Eiffel Project asked students to indicate how much their knowledge had improved in five core content areas as a result of using computers in the school (no improvement, little improvement, considerable improvement, or great improvement; Denes, 1999a). Finally, the students were asked to indicate on a four-point scale (none, a little, quite a bit, or a lot) their responses to a list of items beginning with this stem: "since you started to use computers at school to what extent…"

- Do you come to school more often?
- Do you like school more?
- Do you feel more comfortable writing reports?
- Are you more likely to research colleges on-line?
- Are you more likely to research careers on-line?
- Are you more likely to collaborate with other students on-line doing projects?
- Are you more likely to offer feedback to other students about their work?
- Are you more likely to receive feedback from other students about your work?
- Are you more likely to use computers outside of school?

The San Antonio Technology in Education Coalition addressed a distinct area of learner affect with the School Alienation Scale (Shoho, 1997). This scale is a 40-item instrument that includes the following instructions to students: "below are some statements regarding your perceptions about school with which some people agree and others disagree. Please circle the answer that best expresses your feeling." The survey makes use of a standard five-point Likert scale. The survey addresses feelings of competence, social belonging, and attitudes toward school held by the participant as well as the participant's estimation of her peers' attitudes. Sample items include:

- It is all right to break school rules as long as you do not get caught.
- I feel uncomfortable at school.
- What I'm learning at school is not helping me prepare for life after school.
- I don't care what kind of grades I earn.
- I have a difficult time interacting with my classmates.
- Succeeding in school is important to me.
- I feel so small at this school.
- Students don't have any control over what happens at this school.
- My friends don't think school means anything.

This measure allows several dimensions of learner affect to be examined, all within the context of belongingness versus isolation. This scale is unique in providing an estimation of locus of control (i.e., "students don't have any control"), which ties into learner motivation through attributions for success and failure (Weiner, 1994).

Ratings of Affect by Significant Others

The second general approach to evaluating learner affect is using significant others' ratings of interest, attitude, and perceptions of competence. The two primary sources for these data are parents and teachers. In most projects teachers were asked to rate how their entire class was affected by a project; they did not rate the response of individual students. Parents were asked to rate the response of their own child.

The **Eiffel Project** augmented self-report data with a teacher questionnaire that addresses several topics, including learner affect. Teachers were asked to "indicate your expectations about the extent to which participating in the Eiffel Project is likely to affect your students' school-related attitudes and behaviors." The list of items to which the teacher responds includes items focusing on both cognitive and affective outcomes. The rating scale for this section of the questionnaire asks the teacher to choose among: none, a little, somewhat, and a great deal. In addition to the following affective outcome items, there is an additional line for the teacher to add comments under the heading "other" (Denes, 1999b).

- Increase liking of school
- Increase confidence as a learner
- Increase interest in technology
- Increase class participation
- Improve student-student cooperation

The **Teacher Led Technology Challenge** uses a similar procedure to the Eiffel Project, by embedding items regarding learner affect into a teacher questionnaire, but adds another level of decision making for the teachers (Berkeley Planning Associates, 1999a). In the TLTC questionnaire, the teacher is asked to not only make a rating regarding the affective outcome, but to make a differential response for the experimental group (TechnoKids) and the comparison group (other students). Furthermore, the teachers are asked to identify where they gathered the information that led to their ratings, choosing "all that apply" from the following list:

- Observation of students in classroom
- Performance on class assignments and tests
- Feedback from parents
- Reading assessments
- Standardized test scores
- Other _________________

As for the decision points, the teacher provides ratings for both cognitive and affective outcomes, with the directions asking them to "check the response that indicates how much difference participating in the TLTC project has made for your students." The levels of rating provided by the scale are: a great deal, a fair amount, a little, and none. Thus, the teacher provides estimates for his TechnoKids students and other students to the following items targeting learner affect:

- Increased liking of school
- Improved student-student cooperation
- Improved ability to work independently
- Increased leadership skills
- Increased self-confidence

The results gathered from this instrument indicated that teachers judged the majority of students benefited at least "a fair amount" in two of the affective areas: liking of school and student-student cooperation (Toms Barker, Weinstock, & Gearhart, 1999).

In addition to the teacher questionnaire, the TLTC project also makes use of parental reports to address learner affect (Berkeley Planning Associates, 1999b). The parent questionnaire addressed several aspects of the parents' attitudes regarding the project, but in particular addressed select aspects of learner affect. One method of addressing learner affect used with this scale was to identify activity choices regarding the computers associated with the project. The parents were asked to indicate how much time the child spent with the computer, who the child worked with when working on the computer, and the types of activities used by the child with the computer. To assess chosen activities, the parents were asked to select the activities the "child chose" while she "worked on the computer."

– Simply chose him/herself from the software available on or with the computer.
– Your child's teacher assigned or suggested that your child work on.
– You or another adult in the household suggested that your child work on.
– Other (please explain): ______________________

Parents also reported on affective dimensions similar to the teacher report, allowing validation of ratings. The parent/guardian instrument asked "how satisfied are you with" several topic areas, with the following options: very dissatisfied, dissatisfied, satisfied, and very satisfied. Affective items in this scale were:

– Ability to cooperate with others
– Ability to concentrate on a task
– Motivation for learning
– Self-confidence

Finally, the parent or guardian was provided with open-ended items that could have solicited affective results, including "what do you think that your child has gained from participation in the TechnoKid Program?" The results from the parent ratings and open-ended items allowed for several statements regarding improvements in the affective domain, including increased levels of satisfaction for the child's motivation for learning from baseline to follow-up. Several parents also indicated in the open-ended section that the children had "developed a love — or an increased interest—for learning through working on the computer (Toms Barker, et al. 1999)."

An important part of the intervention in the **Delaware Challenge** project was having children take home Sony Play Stations with Lightspan software that they could use to practice reading concepts introduced in the classroom. Parents were asked in a survey to indicate how “the amount of time devoted to specific activities that your child participates in has changed” since their participation in the project. The parents were asked to indicate whether amount of time in activities “increased, stayed the same, or decreased” for watching television or videos, schoolwork, playtime, and activities with family. These ratings infer affect using the indicator of activity choice, which is driven largely by motivation and interest. Parents were also asked to rate the statement, “my child enjoys using the Lightspan CD’s” on a five-point scale: strongly disagree, disagree, agree, strongly agree, and don’t know (Delaware Education Research and Development Center, 1999).

Classroom Observations

In addition to surveys, external reports on learner affect are available through archival searches and systematic observations. Observations vary widely in the format and focus, dependent upon the project, thus selected styles will be presented.

The **Teacher Led Technology Challenge** made use of an involved observation tool that addressed roles of teachers and students, and included a location for a sketch of the classroom. The observation form also calls for information regarding the contextual setting of the observation, including number of students, type and order of activities, intention of the content presented in the lesson observed, and the presence and structure of student groups. The primary function of this observation measure was to describe the climate and activity during the observation. With respect to affect, the primary focus was on behavior. In particular, a set of behavioral ratings identified if students participated in group-interaction behaviors, and if so, the rater further indicated if the interaction had “substance” or made use of “technology.” The behaviors noted here were: task relevant conversations, describing and sharing, and assisting others. The observation also asked for an evaluation of the class’ overall behavior pattern regarding “behaving appropriately to the task” and staying “focused and attentive.” The observer indicated on these items that either: most, some, or few or none. This rating was followed by an explanation for the rating, and an explanation of the teacher’s response (Berkeley Planning Associates, 1997).

The **A.C.T. Now! (Anderson, IN)** project makes use of a classroom observation form as well to identify various project impacts, including learner affect as related to the project goals. The affective items on this classroom observation tool (which requires 20 to 30 minutes of observation time overall to complete) included three basic items that were rated on a four-point scale: always, usually, sometimes, never (Cassady, 2000). The items were:

- Students were comfortable with technology usage
- Students tried several things pertaining to technology without teacher support
- Students were willing to help each other with technology-centered activities

The results from this scale provided results that indicated higher levels of comfort and self-confidence with technology related activities (Cassady & Cassady, 2000). These results were possible through observations of both experimental and comparison classrooms, with the additional factor of grade level providing significant influence to the results. The developmental factor is important for several projects such as the A.C.T. Now! project, due to the confound existing between years in the project and age or grade level, as the project builds upon experiences over time in the schools.

Archival Searches

Archival searches refers to using the collection of student performance data designed for other purposes and available in some form of institutional database. One example of the use of archival data is the **A.C.T. Now! (Anderson, IN)** project. The measures of learner affect include:

- Attendance rates of students in project versus comparison schools.
- Behavioral referral rates of students in project classrooms versus comparison classrooms.
- Class enrollment and selection for elective courses involving technology
- Participation in after-school activities related to academics or technology
- Involvement in parent-child activities associated with the school
- Use of project-related materials in local public library

The systematic approach to gathering these data included e-mail, fax, mail, and phone calls to the principals and administrators in the school district. The process was aided by the presence of a district-level database system that tracks grades, referrals, standardized test scores, and attendance for several years. However, some data needed to be gathered from student folders or school summary reports.

Commentary

Overall, information on learner affect is strong in the TICG and Star Schools programs. Evaluators seem to have embraced the notion that these data are important, and provide relevant information to program efficacy. Part of the impetus behind the high visibility of data on affect may be the interest of the stakeholders, as there is generally great interest in students' emotions, motivations, and beliefs. Although the instruments being used to gather data on learner affect appear to be strong in most cases, the overall quality of information available could be strengthened if more projects made use of established measures. The continual re-creation of instruments for each project is not only time-consuming, it makes the comparison of projects difficult. Naturally, there would be situational adaptations across projects to several items, but the development of a core set of items would enhance the ability to gather meaningful data allowing comparative analyses of the impact of projects on learners' attitudes, motivation, and beliefs. Furthermore, the use of instruments across several settings would allow for higher confidence in the validity of results and conclusions due to the availability of psychometric data enabled by larger sample sizes.

Finally, it appears that several projects are models for making use of varied methodologies for identifying affect. All projects would benefit from this multi-method approach to examining the affective constructs of interest to their stakeholders. Clearly, there are ample resources available for all methods of data collection, however the process of disseminating the evaluation tools is only beginning to develop. Ideally, collaboration among evaluators will continue, leading to a more complete set of tools for assessing learner affect through varied methodologies.

References

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84,* 261-271.

Bandura, A. (1989). Human agency in social cognitive theory. *American - Psychologist, 44,*1175-1184.

Berkeley Planning Associates (1997). *Classroom observation summary: Elementary and middle schools.* (Teacher Led Technology Challenge.) Oakland, CA: Author.

Berkeley Planning Associates (1998). *Technology skills checklist: Grades 6-8.* (Teacher Led Technology Challenge.) Oakland, CA: Author.

Berkeley Planning Associates (1999a). *Follow-up teacher questionnaire: Elementary and middle schools.* (Teacher Led Technology Challenge.) Oakland, CA: Author.

Berkeley Planning Associates (1999b). *Parent/guardian follow-up questionnaire: Elementary and middle schools.* (Teacher Led Technology Challenge.) Oakland, CA: Author.

Berndt, T. J., & Keefe, K. (1995). Friends' influence on adolescents' adjustment to school. *Child Development, 66,* 1312-1329.

Brophy, J. (1999). Toward a model of the value aspects of motivation in education: Developing appreciation for particular learning domains and activities. *Educational Psychologist*, *34,* 75-86.

Cassady, J. C. (2000). *The classroom observation form.* (A.C.T. Now! Project.) Muncie, IN: Ball State University.

Cassady, M. D., & Cassady, J. C. (2000). *Spring 2000 Classroom Observation Preliminary Report.* (A.C.T. Now! Project.) (Submitted to Anderson Community Schools Corporation) Muncie, IN: Ball State University.

Cassady, J.C., Mantzicopoulos, P.Y., & Johnson, R.E. (August, 1996). *The Accuracy and Multidimensionality of First and Second Grade Students' Academic Self-Concepts.* Paper presented at the American Educational Research Association for the 1997 Annual Meeting, Chicago, IL.

Csikszentmihalyi, M., & Nakamura, J. (1989). The dynamics of intrinsic motivation: A study of adolescents. In C. Ames & R. Ames (Eds.), *Research on motivation in education.* San Diego, CA: Academic Press.

Delaware Education Research and Development Center (1999). *Delaware Challenge: Year four evaluation report.* Newark, DE: University of Delaware.

Denes, V. (1999a). *Eiffel Project student survey.* New York: Columbia University, Teachers College, NCREST.

Denes, V. (1999b). *Eiffel Project teacher questionnaire.* New York: Columbia University, Teachers College, NCREST.

Dweck, C. S. (1986). Motivational processes affective learning. *American Psychologist, 41,* 1040-1048.

Entwisle, D.R., Alexander, K.L., Pallas, A.M., & Cadigan, D. (1987). The emergent academic self-image of first graders: It's response to social structure. *Child Development, 58,* 1190-1206.

Ertmer, P. A., Evenbeck, E., Cennamo, K. S., & Lehman, J. D. (1994). Enhancing self-efficacy for computer technologies through the use of positive classroom experiences. *Educational Technology Research and Development, 42*(3), 45-62.

Fetterman, D. M. (1988). *Qualitative approaches to evaluation in education: The silent scientific revolution.* New York: Praeger.

Furman, W., & Buhrmester, D. (1992). Age and sex differences in perceptions of networks of personal relationships. *Child Development, 63*, 103-115.

Graham, S., & Weiner, B. (1996). Theories and principles of motivation. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 63-84). New York: Macmillan Library Reference.

Harter, S. (1982). The Perceived Competence Scale for Children. *Child Development, 53*, 87-97.

Harter, S. (1983). Developmental perspectives on the self-system. In P.H. Mussen (Gen. Ed.), E.M. Hetherington (Vol. Ed.), *Handbook of child psychology*, (Vol. 4; pp. 275-385). New York: Wiley.

Harter, S., & Pike, R. (1984). The pictorial scale of perceived competence and social acceptance for young children. *Child Development, 55,* 1969-1982.

Helmke, A., & Van Aken, M.A.G. (1995). The causal ordering of academic achievement and self-concept of ability during elementary school: A longitudinal study. *Journal of Educational Psychology, 87,* 624-637.

Henerson, M., Morris, L.L., & Fitz-Gibbon, C.T. (1987). *How to measure attitudes* (2nd ed.). Newbury Park, CA: Sage.

Herrnstein, R.J. (1977). The evolution of behaviorism. *American Psychologist, 32,* 593-603.

Horowitz, J., & Brown J.H. (1996). Confluent education and evaluation research. In *Advances in confluent education,* 113-141. JAI Press.

Hull, C. (1951). *Essentials of behavior.* New Haven, CT: Yale University Press.

Judd, C.M. (1987). Combining process and outcome evaluation. In M.M. Mark & R.L. Shotland (Eds.), *Multiple methods in program evaluation* (pp. 23-42). San Francisco: Jossey-Bass.

Kahne, J. (1996). The politics of self-esteem. *American Educational Research Journal, 33,* 3-22.

Kirshstein, R., Lovitts, B., Lundmark, V., & Weidler, D. (1999). *The 1995 technology innovation challenge grant projects: highlights from the interim reviews (draft).* Washington, D.C.: American Institutes for Research.

Kidder, L.H., & Fine, M. (1987). Qualitative and quantitative methods: When stories converge. In M.M. Mark and R.L. Shotland (Eds.), *Multiple methods in program evaluation* (pp. 57-76). San Francisco: Jossey-Bass.

Lynch, E. W. (1999). *ACT Now! student technology survey.* (ACT Now! Project.) San Diego: San Diego State University.

Lynch, E. W., Bober, M., Harrison, P. J., Richardson, H., & Levine, S. (1998). *ACT Now! case study instruments.* San Diego: CA: San Diego State University.

Marchant, G. J., Paulson, S. E., & Rothlisberg, B. A. (in press). Relations of middle school students' perceptions of family and school contexts with academic achievement. *Psychology in the Schools.*

Marsh, H.W. (1989). Age and sex effects in multiple dimensions of self-concept: Preadolescence to early adulthood. *Journal of Educational Psychology, 82,* 646-663.

Marsh, H.W. (1990a). Causal ordering of academic self-concept and academic achievements: A multiwave, longitudinal panel analysis. *Journal of Educational Psychology, 82,* 646-663.

Marsh, H.W. (1990b). A multidimensional, hierarchical model of self-concept: Theoretical and empirical justification. *Educational Psychology Review, 2* (2), 77-172.

Marsh, H. W., Barnes, J., & Hocevar, D. (1985). Self-other agreement on multidimensional self-concept ratings: Factor analysis and multitrait-multimethod analysis. *Journal of Personality and Social Psychology, 49,* 1360-1377.

Marsh, H.W., & Craven, R.G. (1991). Self-other agreement on multiple dimensions of preadolescent self-concept: Inferences by teachers, mothers, and fathers. *Journal of Educational Psychology,83,* 393-404.

Marsh, H.W., & Shavelson, R.J. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist, 20,* 107-125.

Maruyama, G. (1998). *Basics of structural equation modeling*. Thousand Oaks, CA: Sage Publications.

Merker & Associates (1999a). *Student survey.* Mississippi Educational Television Interactive Video Network. Louisville, KY: Merker & Associates.

Merker & Associates (1999b). *1999 METIVN Star Schools project evaluation: Report 3.* Louisville, KY: Merker & Associates.

Mucherah, W. (2000). *Student computer use survey.* Baltimore: Baltimore Learning Community/Maryland Early Learning Community.

Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review, 91,* 328-346.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory (3rd ed.).* New York: McGraw-Hill.

Popham, W.J. (1988). *Educational evaluation* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Rawson, H.E., & Cassady, J.C. (1995). Effects of therapeutic intervention on self-concepts of children with learning disabilities. *Child and Adolescent Social Work Journal, 12,* 19-31.

Research and Training Associates (1999). *Evaluation of the Trails Project.* Overland Park, KS: Author.

Sawyer, R.J., Graham, S., & Harris, K.R. (1992). Direct teaching, strategy instruction, and strategy instruction with explicit self-regulation: Effects on the composition skills and self-efficacy of students with learning disabilities. *Journal of Educational Psychology, 84,* 340-352.

Schunk, D.H. (1990). Self-concept and school achievement. In C. Rogers & P. Kutnick (Eds.), *The social psychology of the primary school* (pp. 70-91). London: Routledge.

Schunk, D. H. (1998). Teaching elementary students to self-regulate practice of mathematical skills with modeling. In D. H. Schunk & B. J. Zimmerman (Eds.) *Self-regulated learning: From teaching to self-reflective practice.* (pp. 137-159). New York: Guilford Press.

Schunk, D.H. (1999). Social-self interaction and achievement behavior. *Educational Psychologist, 34,* 219-227.

Shavelson, R.J., Hubner, J.J., & Stanton, G.C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46,* 407-441.

Shoho, A. R. (1997). *School alienation scale.* (San Antonio Technology in Education Coalition.) San Antonio: University of Texas at San Antonio.

Shotland, R.L., & Mark M.M. (1987). Improving inferences from multiple methods. In M.M. Mark & R.L. Shotland (Eds.), *Multiple methods in program evaluation* (pp. 77-94). San Francisco: Jossey-Bass.

Snow, R.E., Corno, L. & Jackson, D. (1996). Individual differences in affective and conative differences. In D. C. Berliner & R. C. Calfee (Eds.) *Handbook of educational psychology* (pp. 243-310). New York: Macmillan Library Reference.

Stipek, D., & MacIver, D. (1989). Developmental change in children's assessment of intellectual competence. *Child Development, 60,* 521-538.

Stipek,D., Recchia, S., & McClintic, S. (1992). Self-evaluation in young children. *Monographs of the Society for Research in Child Development, 57*(1).

Stone, N. J. (2000). Exploring the relationship between calibration and self-regulated learning. *Educational Psychology Review, 12,* 437-475.

Toms Barker, L., Weinstock, P. & Gearhart, M. (1999). *Teacher Led Technology Challenge: Year Three Evaluation Report.* (Submitted to Berkeley Unified School District.) Oakland, CA: Berkeley Planning Associates.

Toms Barker, L., Weinstock, P., & Frakes, D. (2000). *Teacher Led Technology Challenge: Year Four Evaluation Report.* Oakland, CA: Berkeley Policy Associates.

Weiner, B. (1994). Integrating social and personal theories of achievement striving. *Review of Educational Research, 64,* 557-574.

West, S. G., Biesanz, J. C., & Pitts, S. C. (2000). Causal inference and generalization in field settings: Experimental and quasi-experimental designs. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology.* (pp. 40-84). Cambridge, United Kingdom: Cambridge University Press.

Wigfield, A. (1997). Reading motives: A domain-specific approach to motivation. *Educational Psychologist, 32,* 59-68.

Wigfield, A., Eccles, J. S., & Rodriguez, D. (1998). The development of children's motivation in school contexts. *Review of Research in Education, 23,* 73-118.

Wylie, R. C. (1979). *The self-concept* (Vol. 2). Lincoln, Nebraska: University of Nebraska Press.

Zimmerman, B.J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal, 29,* 663-676.

3. Learner Outcomes in Adult Education

Jerome Johnston
Institute for Social Research
University of Michigan

Most of the technology projects funded under Star Schools and Technology Innovation Challenge Grants are aimed at K-12 learners. But there are four projects that target adult learners who need a high school credential and/or who need to improve their employment skills so they can qualify for a better job. The products from these projects are designed for use by agencies that serve the educational and job-training needs of a wide variety of adult learners. Each of the projects utilizes technology as a primary instructional and service-delivery tool. Service delivery models range from traditional classroom instruction to independent distance learning.

Evaluation of adult education projects calls for measures that are unique in the educational evaluation field. Cognitive and affective outcomes are valued in adult education, but the goals being assessed are sufficiently different from K-12 to require specialized measures. For example, several projects are aimed at helping adults who failed to earn their high school diploma attain a GED—a certificate of high school equivalency. The various subtests that comprise GED certification are of little interest to K-12 educators. Other projects teach job-seeking skills and workplace reading and math. These topics do not appear on the K-12 agenda except in vocational education. A similar situation exists in the affective arena. Until recently, research in adult education paid little attention to learner motivation. While this is changing, the focus is on motivation to engage in basic literacy activities (reading a newspaper or writing a letter), not motivation to perform better on academic tasks in the classroom. Also, while evidence of cognitive and affective impact is valued, policy makers in adult education are more interested in distal outcomes such as securing or improving a job, or performing better in job settings. Cognitive and affective growth are merely enabling, or intermediate, outcomes for these more valued goals.

Evaluations of new adult education instructional series must also be responsive to the requirements of the new National Reporting System for adult education. Beginning in July of 2000 adult education providers are required to justify the student *per capita* funds they receive by showing that each student makes adequate educational progress while in their program. To promote adoption of new adult education instructional series, evaluation studies should be designed to demonstrate how much educational progress a student might be expected to make if they use the new materials.

This chapter begins with a brief description of five adult education projects. It then describes the National Reporting System and its classification system of adult learners. Then the chapter provides examples of the types of measures being used to evaluate three of the five programs.

Technology-Based Adult Education Projects

Four adult education projects were funded by the Department of Education's Star Schools program. They are designed primarily for adults who are: over age 18, lack a high school diploma or GED, are unemployed or hold a job where the pay and benefits are below their personal economic needs, possess job skills that are insufficient to obtain better employment, and are either native English speakers or non-native speakers who comprehend English well enough to engage the educational materials.

- **Workplace Essential Skills (WES)** (broadcast video, video cassette, print, online) from the PBS LiteracyLink project. WES teaches strategies for finding and succeeding in a job. It also covers the many ways that communication, reading and math are used on the job. Twenty-four units released Fall, 2000.

- **GED Connection** (broadcast video, video cassette, print, online) also from the PBS LiteracyLink project. Review courses designed to prepare students for each of the five subtests associated with the new GED test scheduled for release in 2002. The subtests cover writing, social studies, science, interpreting literature and the arts, and mathematics. Forty units released Fall, 2001.

- **ESD 101—Educational Service District 101** (broadcast video, print) from the Step*Star Network/Adult Education-Literacy Project. Programs designed to help adults earn a high school credential, improve their English skills, and/or develop "workplace reading skills." Distance education using live studio teachers and toll-free lines for tutorial support. Released Fall, 1996.

- **CLASS Project** (online) from the University of Nebraska. A collection of high school completion courses meant to be studied at home. Students can earn a high school diploma from the state of Nebraska. About half of the students enrolled are aged 16-19. This program does not have the typical profile of programs designed especially for adults. Courses released continually beginning in Fall, 1998.

Further information on these projects can be found at the Star Schools Web site (Star Schools, 2001).

A fifth program—TV411—is discussed in this chapter because it has a complementary relationship with several of the Star Schools projects. Although the project's materials development efforts are not funded by federal sources, a portion of its evaluation is supported by U. S. Department of Education funds.

- **TV411** (video, print) from ALMA—the Adult Literacy Media Alliance. Promotes interest in reading, writing and everyday math among adult basic learners. Twenty units released since Fall, 1997; additional units in production. Online component under development.

The Target Audience for Adult Education Programs

The broad arena of adult education is known as ABE—Adult Basic Education. Services in ABE are tied to level-based needs of the learners. As part of its new *National Reporting System for Adult Education,* the U. S. Department of Education, Office of Vocational and Adult Education (OVAE), has established six "educational functioning levels" based on reading, math, and communication/language scores as measured by standardized tests (Office of Vocational and Adult Education [OVAE], 2000). These levels are outlined in the table below. The projects covered in this chapter are designed primarily for learners functioning at levels III–VI.

Table 1. NRS Educational Functioning Levels

Level	Level Descriptor	Grade Level Equivalents
I	Beginning ABE Literacy	0 – 1.9
II	Beginning Basic	2.0 – 3.9
III	Low Intermediate Basic	4.0 – 5.9
IV	High Intermediate Basic	6.0 – 8.9
V	Low Advanced Secondary	9.0 – 10.9
VI	High Advanced Secondary	11.0 – 12.9

The majority of adult education programs administer reading, math, and communication (writing and/or oral language) tests as part of student intake and as a means of monitoring student progress in their program. Standardized measures commonly utilized to assess these levels include the Test of Adult Basic Education

(TABE), the Comprehensive Adult Student Assessment System (CASAS), the Adult Measure of Essential Skills (AMES), and the Adult Basic Learning Examination (ABLE).

The NRS has also established six educational functioning levels for non-native speakers, or English-as-a-Second Language (ESL) learners. These range from beginning ESL literacy to high-advanced ESL. Most ABE programs require non-native speakers to demonstrate at least minimal mastery of the English language in the areas of listening, speaking, and reading before they can be admitted to an ABE program of study. If the learner does not, they must study English in an ESL program designed specifically for this purpose. Table 2 shows how the various projects target learners in just a few levels.

Table 2. Target ABE and Reading Levels for the Adult Education Projects

ABE Level	Reading Grade Level	TV411	Workplace Essential Skills	GED Connection	ESD101	CLASS
I	0–1.9					
II	2.0–3.9					
III	4.0–5.9	X	X			
IV	6.0–8.9	X	X		X	
V	9.0–10.9		X	X	X	X
VI	11.0–12.9			X	X	X

Implications for Studies of Adult Learners

Two characteristics of adult learners have important implications for evaluation designs. Given the limited reading proficiency of adult learners, it is often inadvisable to use respondent-completed research instruments. Although they are capable of communicating orally, low-reading level adults lack the skill to decode instructions and written questions in the ways intended by the researchers. They are also limited in their ability to respond clearly in writing. The questionnaire genre (multiple-choice responses, rating scales) is unfamiliar to people with limited formal education. In the TV411 studies the evaluators judged that the average reading level for the target audience was low enough (4.0-8.9 on the TABE) that they needed to use individual face-to-face interviews with all respondents (Johnston, Isler-Petty, and Young 1999, 2000, 2001a). In the WES studies reading levels were higher and teachers (with training provided) were given the task of collecting data from students using self-completed questionnaires. But forms were sometimes filled out incorrectly, data collection procedures were not always adhered to, and assessments and forms were administered inappropriately, raising questions about the validity of the data. The alternative is to have research staff on site for all of the data collections—a prohibitive expense in a large-scale study (Johnston, Young, and Isler-Petty, 2001b).

Other characteristics of adult learners make them less than ideal candidates for research studies with tight time frames. Typically, adult learners drop in and out of school. For example, for their GED preparation program one participating research site expects that 50 percent of the adult students admitted to the program will drop out before the end of the first 10-week period. These dropouts may return to school at some time in the future, but not in time to provide data for a one-semester test of a new educational program.

Unlike K-12 where there is fair degree of uniformity in the organization and delivery of instruction, adult education is conducted by a variety of agencies including community-based non-profit organizations, community colleges, libraries, and prisons. The systems of delivery are less organized than K-12, and many of the instructors are part-time or volunteer. This has implications for using the delivery system to help with the research effort. In K-12 research it is often easy to gain the cooperation of the school and teachers to help with recruitment of subjects, systematic delivery of the intervention to be tested, and routine administration of questionnaires by a staff accustomed to this role. It is more difficult in adult education, even if the institutions and staff are paid for their efforts. The conclusion of this author is that extra resources are needed in adult education studies to insure adequate trained staff are available to assist with all aspects of the research effort: student recruitment and retention, implementation of the intervention under study, and all aspects of data collection.

Learner Outcomes in Adult Education

The learning goals of the projects in this chapter fall into three categories: (1) workplace skill development and job attainment (or job improvement), (2) academic skill building or certification, and (3) orientation to literacy or schooling. Table 3 lists the goal areas for each project and indicates whether the goals were assessed in published evaluation studies.

Historically, workplace skills meant the “hard” skills such as how to operate or repair a machine. Recently, in response to the SCANS report (U.S. Department of Labor, 1992), there is a growing recognition that a set of “soft” skills are needed to secure a job, and to work effectively with other employees once the job has been attained. Job seeking skills include knowing about the employee selection processes, preparing a resume, and presenting oneself in a job interview. Workplace skills include communication, reading, and math skills such as writing a memo, filling out a job ticket, reading data tables, and computing such things as areas and volumes.

Academic skills are those associated with the major academic attainment measure in adult learning: the GED. Literacy orientation refers to the orientation of learners to those activities required to attain the outcomes in the previous two categories—everyday reading, writing, and mathematics. By definition, ABE programs are designed for adults who failed to acquire rudimentary academic and job skills during the normal period for pursuing a K-12 education. Many adults in this category have so little confidence in their own learning skills that they shy away from the very programs that could help them make up their deficits. This group needs to re-focus themselves around basic literacy and learning. They need to reduce their anxiety and increase their motivation for literate activity. Two of the programs—TV411 and WES—have this goal.

Table 3. Project Goal Areas and their Assessment in Evaluation Studies

Learner Outcomes	TV411	WES	GED Connection	ESD101	CLASS
Workplace Skills/Job Attainment					
Job Seeking Skills		G/M		G/M	
Workplace Communication		G/M			
Workplace Reading		G/M			
Workplace Math		G/M			
Employment Status		G/M		G/N	
Academic Skills/Certification					
Course Content Tests			G/N		G/N
High School Diploma					G/N
GED Scores/Attainment			G/N	G/N	
Orientation to Literacy/Schooling					
Knowledge of Basic Literacy Facts	G/M				
Interest in Literacy Activity	G/M				
Confidence in Literacy Skills	G/M				
Anxiety	G/N				
Motivation Towards School	G/M	G/M			

KEY: G = Goal area for the project; M = measured; N = not measured in published studies. In the case of LiteracyLink, the *GED Connection* materials will be completed at the very end of the project, too late to permit evaluating their impact under the initial grant.

Although included in the above table, CLASS and its evaluation look very much like a typical high school program and assessment. The summative evaluation assesses the individual courses in a variety of ways including reviewing the degree to which course objectives reflect skills and content knowledge reasonable for a course at its level, mapping course objectives to the Nebraska State Standards, reviewing the degree to which student assessments reflect course objectives and analyzing student achievement of course objectives. In addition the summative evaluation assesses the course development framework by having experts review the framework and having teachers review the content/pedagogical aspects of the framework and having programmers test the technical aspects of the framework. For this reason, the CLASS project evaluation is not discussed in this chapter (P. Campbell, personal communication, March 26, 2001).

Workplace Skills

Workplace skill development/job attainment programs have the ultimate goal of helping students find a job, advance within an existing job, or secure a better job. Indeed, one of the most valued measures of program success in the NRS system is obtaining employment in the quarter following exit from the program, and maintaining employment for at least three quarters after that. But to improve the likelihood that a student will get a better job, they need to focus on a number of workplace skills demanded by employers. These can be divided into three areas:

1. Workplace behavior such as interacting with coworkers or customers, working on a team, and understanding and meeting employer expectations.
2. Workplace products such as completing a job application or resume, or—on the job—preparing memos, reports, work orders, requests, messages, etc.
3. Workplace literacy development—workplace reading and math.

Assessing workplace skills became more important in the year 2000 when federal reporting standards went into effect for adult education programs. Between 2000 and 2005 a state's federal grant for adult education is tied to demonstrated progress of each adult in a program. States are mandated to show that each student in a program advances one "educational functioning level" during his/her time in class. Each state is given latitude in selecting a test or testing system to measure this gain. (See OVAE, 2000, for a description of the reporting system requirements). An educational functioning level is defined in one of two ways. For standardized tests such as the TABE or CASAS the test developers have established benchmarks for their tests that correspond to different levels of achievement. For non-standardized tests, the benchmarks need to be declared and then defended prior to using the test. States are also permitted to use performance checklists where clusters of increasingly difficult behaviors are divided into 4-6 levels. Teachers can use the checklist to assess competency at intake and again at exit from the program.

Two of the adult education projects—Workplace Essential Skills (WES) and ESD101—have the goal of teaching skills that will improve a person's chances of securing a job and being a productive worker. This section describes briefly some of the measures used in the WES evaluation. Detailed descriptions can be found in the full evaluation report (Johnston, Young, and Isler-Petty, 2001). The ESD101 measures are not yet published.

The evaluation design for the WES evaluation used a repeated-measures design, with no control group, as shown in the diagram below. The study was carried out in 20 sites located throughout the United States.

PRETEST	WES INSTRUCTION	POSTTEST	POST-POST
(knowledge/skill measures immediately prior to instruction)	(6 hours per lesson; up to 25 lessons)	(knowledge/skill measures immediately after instruction)	(self-report on job search & attainment 3 months after instruction)

A content analysis of the WES curriculum was conducted and compared with content coverage of two standardized employment tests: the CASAS and the TABE. Neither of them was perfectly suited to the task. The best match was found with the CASAS Employability Competency System (ECS). The match was reasonably good in the areas of reading, math, and workplace writing. For other WES content areas no existing standardized assessments were found, so tailored knowledge and performance tests were developed for these areas. Table 4 shows all of the outcome measures, showing a mix of standardized measures (all the items with CASAS in the title), tailored performance tasks, and tailored knowledge tests.

Job Seeking Skills

WES teaches "good practices" regarding a variety of job seeking practices (e.g., how to find employment opportunities, what to include in a resume, how to present oneself in a job interview). Assessing mastery of these skills is difficult. While these practices have a knowledge component that can be assessed with a knowledge test, the desired outcomes are much more complex than mere knowledge. The goal is to develop a learner's skills in finding available jobs, applying for them, and presenting themselves to employers in a way that gets them the job. Some of the skills such as completing a job application and writing a resume can be measured with performance tests. But the more subtle skills and behaviors associated with efficient job searching and presenting oneself to an employer require observation of a learner in a real or hypothetical situation and scoring the performance using trained observers. This was not possible in a dispersed-site evaluation design where the research sites were spread widely across the United States. Instead, three tests were used to cover the employment skills area: a standardized performance test for job applications, a tailored performance test for preparing a resume, and a knowledge test of good practices.

Table 4. Summary of WES Measures of Workplace Skills/Job Attainment

Learner Outcome Area	**Measures**
Job Seeking Skills	· CASAS Functional Writing Assessment (fill out a Job Application) · Tailored Resume Task · Tailored knowledge test of good job seeking practices
Job Attainment	· Self Report of Recent Job Seeking Behavior (at the end of class and again 3 months later)
Workplace Communication—oral and written	· Tailored knowledge test of good practices in written, oral, and non-verbal communication in the workplace · Tailored performance tasks: write a memo, fill out an internal office form, and interpret a workplace chart
Workplace Reading	· CASAS ECS Reading Test
Workplace Math	· CASAS ECS Math Test

The CASAS Functional Writing Assessment— A Job Application Task. The CASAS Functional Writing Assessment is designed to measure students' awareness of appropriate information to include in a job application and the appropriate way to present the information on the application. The test consists of filling out a job application. CASAS utilizes a five-point scoring rubric to grade the completed application for content, mechanics (spelling, capitalization, and punctuation), and legibility and appearance. Participants in the WES employment class showed a nine percent increase in their scores, pre to post. This amounts to a one-half level increase in the CASAS five-level system—a respectable improvement given the instructional time frame.

Writing a Resume—A Tailored Performance Task. For this tailored task students were asked to read a short biography of a fictitious job seeker and to identify the information that would be appropriate to include in a resume: The evaluators developed a rubric to use in assessing the resume. Thirty-nine points were allotted for basic information. In addition, four bonus and five penalty points were available. The basic information included four categories of information: work history, skills, education, and references. As an example, 16 points were possible in the work history section: four points for including the section and 4 points for characteristics of each of the three jobs described in the biography—name and location of employer, dates of employment, and description of duties. Bonus and deduction points were awarded for various stylistic features and inappropriate content.

The evaluators made a difficult research design choice for this measure. This tailored task was given only as a posttest; there was no pretest. For a student who had never prepared a resume—virtually everyone in the evaluation sample—it was deemed too difficult and discouraging a task to be completed as part of a one-hour group-administered pretest. It was important not to discourage the learners who had volunteered to participate in the study. Participants averaged scores of 30 points out of a possible 43. This was judged to be at a level where the resume had the basic elements, but needed additional editorial work before it could be submitted to an employer.

Getting a Job: What's Important—A Knowledge Test. This assessment is a true-false test with 60 items that measure a student's knowledge of good practices for finding a job, completing job applications, creating resumes and cover letters, and interviewing. It also covers a few basic practices regarding learning on the job, appropriate job behaviors, and workplace safety procedures. The items measure the practices advocated in the WES materials. Reliability data are not available at the time of this writing. A sampling of items is shown below.

– Employers almost always run an ad in a newspaper when they have a job available.
– *Transferable skills* are skills that you can use only in a specific job situation.
– It is against the law for an employer to ask the race or color of an applicant on an application form.
– In general, never discuss vacation time, sick days or benefits at a first interview, unless the interviewer brings it up.
– Whenever you send a resume through the mail you should send a cover letter with it.
– If you are confused by your supervisor's instructions, it is better to guess at what your supervisor wants you to do than to bother him or her by asking questions.

On average, participants increased their scores six percentage points, though they knew much of the information prior to the instruction (pretest 76%; posttest 82%). In an exemplary site, the effects were much bigger—the scores increased from 77% to 91%, illustrating how important good teaching is to achieving the learning goals of a program.

Job Attainment

While the instructional components of the job-seeking curriculum teach skills, the ultimate goal is to have the students seek a first-time job or an improved job. In the WES evaluation this was assessed in two ways: once using pre-post measures of job-seeking behaviors and plans; the other with a three-month follow-up interview on the telephone.

Job Attainment Three Months After the Program. A job follow-up survey was conducted over the phone. Students were called at the end of the first quarter following completion of the course. The timing and questions for this interview were those recommended by the NRS for evaluating progress of adult learners. The questions in the brief interview concerned education and job attainment since completing the WES class. Methodologically, the results were quite discouraging. Only 50% of the respondents could be reached and only 32% provided usable data. The most common reasons for not reaching respondents were that the telephone had been disconnected or no one answered the phone in three attempts. Based on conversations with those who did reply, caller-ID may have kept some respondents from picking up the phone. All of the calls were initiated from a location distant from the respondent. Of those who did reply, many were unable to recall that they had participated in the WES field test! Local programs may find it easier to follow up on their own students if they come from areas in the immediate vicinity of the school.

Job Seeking Behaviors and Plans. A more reasonable outcome for an educational program is that students will engage in job-seeking behaviors; being successful at getting a job is subject to many factors beyond schooling. To measure job-seeking behaviors a collection of activities were chosen to sample the job seeking domain. At both pretest and posttest students were twice presented with the list. The first time they were asked, "How often have you done each of the following?" The response scale was (1) Never, (2) Once, (3) More Than Once. The second time they were asked, "How likely is it that you will do each of these things in the next month?" The response scale was (1) Not at all Likely, (2) A Small Chance, (3) Very Likely, and (4) Definitely. Activities in the list included:

– Read the classified ads in a newspaper to look for a job
– Went into a store or business that had a "Help Wanted" sign posted and asked about the job
– Filled out a job application
– Wrote (or revised) your resume

Workplace Communication

A collection of tailored knowledge and performance tests was designed to assess attainment of the workplace communication skills taught in this strand of the WES curriculum. As with the employment strand, a knowledge test was designed to assess skills where a performance measure would be too difficult to implement in a large-scale dispersed evaluation.

- Form Completion task: using information from a memo to complete a room request form.
- Memo task: using information about an upcoming fire drill as the basis for writing a memo to workers in the affected division of the company.
- Chart task: using information from a chart to answer questions.
- True/False questions that assess students' knowledge about types of business communication (e.g., memos, business letters, reports, forms, and other work-related writing tasks), formatting, terminology, form completion, and chart reading. Similar set of questions on verbal and nonverbal communication (e.g., body language, customer relations, presentations, team work, and problem solving).

Increases on the knowledge tests were 11% for the verbal/non-verbal communication and 20% on the written communication, showing that intervention did teach participants to identify important principles. Increases on the performance tests were more varied. Participants showed little to no improvement on the form completion and chart tasks, but they showed a 13% increase on the memo task. It is unclear whether the poor showing on the two tasks was due to weak instructional materials or weak teaching.

Workplace Reading and Math

CASAS tests were chosen for assessing reading and math skills both because of the fit with the WES curriculum and because CASAS is the primary assessment tool used in the adult education reporting systems in many states. Both tests have been normed using data from adult learners in multiple educational and social service settings.

The reading test requires reading and interpreting a variety of text samples—letters, instructions, resumes, tables, and maps. The math test requires reading and interpreting a variety of charts, tables, and instruments. Some questions require identifying the appropriate information in a chart; other questions require completing computations or solving word problems.

The tests come in four forms to accommodate varying ability levels. Prior to administering either of these tests, the ECS Appraisal Form 130 (similar to the TABE Locator Test) was administered and used to determine which form of the full-length reading and math test to use.

CASAS provides a benchmark for expected gains on the Reading and Math tests. After 100 hours of instruction, student scores should improve by five points. In the WES evaluation the increase in reading was two points, but there was only 24 hours of instruction. Extrapolating, it was judged that the WES reading units were sound and would lead to valued gains in adult learners. There were similar findings for math. The average gain for students in the math strand was two points after only 30 hours of instruction.

Because the match between the WES curricula in reading and math and the CASAS tests was not perfect, a secondary analysis was done, eliminating items from both tests that were not directly taught in the WES curriculum. There were no significant differences in the results using the abridged test.

Academic Skills/Certification—the GED

It is a widely held belief by many adults who do not have a high school diploma that securing their GED will give them access to good jobs. This perception is not without foundation; according to the GED Testing Service (developer and distributor of the GED) more than 95 percent of employers in the U.S. consider GED graduates the same as traditional high school graduates with regard to hiring, salary, and opportunity for advancement. That having been said, the current GED tests do not measure many job-related skills that are valued by employers. In an attempt to respond to this criticism the GED Testing Service will use a new set of tests beginning in January, 2002. Success on these tests is supposed to require more application of knowledge from workplace settings than is the case with the tests they replace. When the new tests are released, students who have not yet attained their GED, but who have passed some of the five tests, must start over again accumulating passing scores on the five new tests.

In anticipation of the new test, PBS LiteracyLink developed *GED Connection* to prepare adults for the new test. The 38-part instructional programs (online, print and video) were released in August, 2001—five months before the new GED becomes the new standard for GED certification.

Because *GED Connection* materials were not ready until the project grant was complete, and the most appropriate assessment was not yet released, summative assessment was not included in the grant. Should additional funds be found for evaluation, the study will compare those who take the course with matched samples in other GED preparation programs, using the actual GED test as the outcome measure, with additional data collected on learner performance in the class and time spent studying the *GED Connection* curriculum. Unlike the other adult education programs, the close match between the curriculum and an existing standardized test will greatly simplify assessment. Even though there are no published evaluation studies of technology curricula that prepare for the GED, a description of the test is included in this chapter because the test plays such a central role in adult education.

The term GED stands for General Educational Development. To earn a GED, an individual must attain a passing score on five tests: social studies, science, literature and arts, mathematics, and writing skills—a test with two subparts: correct writing and essay composition. The tests can be taken at different times and taken repeatedly until five tests have been passed. The test descriptions below are minor modifications of the descriptions that GED Testing Service provides prospective test-takers on their website (GED Testing Service, 2001)

Writing Skills. The Writing Skills Test is divided into two parts: a multiple-choice fifty-five question section (Part I), and an essay (Part II). In Part I, test-takers are asked to correct and edit sentences without changing their intended meaning. Part I makes up about two-thirds of the combined score for the Writing Skills Test. In Part II, test-takers compose a response to a question about a subject or an issue. Each essay is scored independently by two trained and certified professionals who grade the essay according to detailed rubrics for effectiveness, clarity, making the main point, and supporting ideas.

Social Studies. The Social Studies Test measures ability to use knowledge and information about a variety of topics in the five subject areas: history, economics, political science, geography, and behavioral sciences (psychology, sociology, and anthropology). Most of the questions in the Social Studies Test refer to information provided in the test itself. The information may be a paragraph, or it may be a chart, table, graph, map, cartoon, or figure. The test-taker must understand, use, analyze or evaluate the information provided to correctly answer the questions.

Science. The Science Test measures skills in problem solving and reasoning as applied in the areas of biology, earth science, physics, and chemistry. Test questions relate to major concepts that cover all areas of science, including change, conservation of mass and energy, interactions, relationships, and time. About half of the questions in the Science Test are based on concepts from biology. As with Social Studies, the Science test requires a person to use graphs, tables, charts, maps and figures to gather the information needed to answer question. Answering correctly requires a person to understand, use, analyze or evaluate the information provided to answer the questions correctly.

Interpreting Literature and the Arts. The Interpreting Literature and the Arts Test measures ability to understand and analyze a text passage. Test questions are drawn from popular literature, classical literature, and commentary excerpts. Many literature selections are taken from works done by American writers, although Canadian, English, and translated pieces from authors around the world are also included. There are excerpts from fiction, non-fiction, poetry, and plays as well as commentary about literature and the arts. The Spanish-language GED Tests use Spanish literature, with an emphasis on Caribbean, Central and Southern American authors. Similarly, the French-language GED Tests draw on French culture.

Mathematics. The Mathematics Test measures ability to solve—or find the best method to solve—mathematical problems typical of those studied in high school mathematics courses. Test questions are drawn from three major areas: arithmetic (measurement, number relationships, data analysis), algebra, and geometry. Questions in the Mathematics section of Spanish- and French-language GED versions are the same as or similar to questions found in the English version, except that most measurements are given using the metric system.

Orientation to Literacy/Schooling

The previous interventions all assume that a learner is aware of his/her deficiencies and inclined to remediate them to improve their job prospects. Many adult learners are not even aware of their deficiencies. Or, if they are aware, they fear they cannot pass a course designed to remediate the deficiency. Earlier failures in school settings have left many believing they are incapable of succeeding in a typical school-based course that has demanding content and assessment. For many of these adults, everyday literacy tasks such as reading food labels, writing a letter, or setting up a budget are challenging. These adults need an "attitude adjustment" course where they can be exposed to a wide variety of everyday literacy activities, encouraged to engage in them, and can experience success when they do engage in them.

TV411 from the Adult Literacy Media Alliance has just this goal. WES has similar goals, but they are embedded in a workplace training course. TV411 is designed to introduce students to a variety of literacy activities and provide them with some direct instruction and information about how to engage such activities. The TV411 materials contain a mix of motivational messages and direct instruction. The television shows contain engaging stories about adults who have turned around their lives. Many of these adults have learned how to read and write late in life and have had to overcome many obstacles to better their lives in this way. These learners' stories, by themselves, have the potential to affect the beliefs and attitudes of viewers regarding the value of engaging in literacy activities and the expectation that they might succeed in doing so.

To measure TV411's effectiveness, evaluators developed measures in two arenas: interest in doing the modeled literacy activities and confidence regarding one's skill in performing the activities. They also assessed knowledge gain using tests tailored to the most prominent skills and factual material in the series. Using a pre-post design, they measured changes in these arenas over the intervention period—eight weeks in one study and ten weeks in another (Johnston, et al. 1999, 2000, 2001a).

Interest in Literacy Activities

A major goal of TV411 is to stimulate the audience to engage in a variety of literacy activities that may not already be part of their daily repertoire. If the participants responded to the suggestions in the shows and to the intrinsic reward they received from engaging in the activities, then there should be a change in their daily repertoire by the end of the test period. To assess this, participants were asked before the test period began and again afterwards about their *expectations* for engaging in each of 23 literacy activities promoted on the TV411 shows included in this test. While expectations to behave and actual behaviors are not the same, they are a reasonable proxy for what people will do and a good indicator of this type of impact. Behavioral expectations can also be viewed as an attitude measure: if behavioral expectations increased over the eight weeks, then the experience has led to a more positive attitude and to a *predisposition* to behave in new ways.

The baseline and follow-up interviews asked about literacy activities in two separate questions. The first question asked: "In life we have routines; we do many of the same things from week to week. I'm going to list a number of activities—like reading a newspaper or writing a letter—and I want you to tell me how often you did this in the last week." There were three choices: (1) No-I didn't do this in the last week, (2) I did it once, or (3) I did it more than once. The second question used the same list and asked: "For each activity I also want you to indicate how likely it is you will do this in the next week." There were four choices: (1) Not at all likely, (2) A small chance, (3) Very likely, and (4) Definitely. The questions were asked in a face-to-face interview, and participants were trained in the meaning of the scales before being presented with the actual items. Sample activities are shown below. The complete list of 23 items can be found in Johnston, Isler-Petty, and Young (2001a).

- Read a newspaper
- Write in a personal diary or journal
- Read a book because the story is interesting to you
- Write a letter to a friend
- Write a letter to a company asking them do something, explaining about something you did, or asking for information
- Write an essay—a short paper—perhaps to describe an experience you had
- Edit a piece of your own writing to make the words or sentences sound better

The assessment detected change in the course of the intervention. At the end of the intervention participants reported a higher likelihood that they would engage in many of the literacy activities in the coming week or month. This was especially true for literacy activities that had been very infrequent in their lives before the intervention.

Confidence in Literacy Skills

A content analysis of the eight shows and workbooks used in the research project identified 20 literacy skills that were modeled in the materials. It was reasonable that participants might have learned — from the messages in the video or from their own successes in the workbooks — that the skills were ones that they could perform better than they had previously thought. In other words, the TV411 experience could have affected participants' confidence in their ability to use these skills.

To measure confidence in these same skill areas participants were asked — both before and after the test period — to rate their confidence in performing each of the activities. The list of activities was adjusted slightly to reflect the differences between simply doing an activity and confidence in performing the activity. Each activity in the list was read aloud during the face-to-face interview and participants were asked to respond on a four-point scale: (1) Not At All Confident, (2) A Little Confident, (3) Pretty Confident, and (4) Very Confident. As with the literacy behaviors, participants were trained in how to use the rating scale. A sampling of the 20 skills appears below.

- Write a personal letter to a friend or relative
- Recognize correct grammar such as where to put an apostrophe, how to make a singular word plural, how to make the subject and verb of a sentence agree
- Edit or rewrite a piece of your own writing to make the words and sentences sound better

- Figure out the price of an item that is on sale—for example, calculate the cost of a $540 television if it is on sale for 30% off

After responding to each of the 20 individual literacy activities, learners were asked three questions about their overall confidence. "As a result of spending the last eight weeks in the TV411 class, would you say you are more or less confident about your ability to read things you come across? …write clearly? …figure out an average or a percentage?" The response scale was (1) Less Confident, (2) About the same, and (3) More confident.

The evaluators found that the confidence measures were sensitive to the experience. All participants showed increased confidence in their literacy skills. Also, there were higher levels of confidence for those that engaged the materials in a facilitated, face-to-face support group built around the TV411 materials than for those who merely studied the materials at home (Johnston et al., 2000, 2001a).

Impact on Educational Behavior

A more stringent test of the intervention was conducted in one test site with very encouraging results. This site offers a GED preparation program that runs for three 10-week semesters. Typically, 50 percent of the students who begin the GED program drop out before the end of the first semester. In Winter of 2001 a group of 36 applicants were identified who fell just short of the minimum requirements on the entry reading test. The school's dean made them this offer: if they participated in an intensive version of the TV411 support group (four 3-hour sessions per week for three weeks) they would be admitted to the regular program despite their low reading score. All the applicants accepted the offer, and all but one completed the TV411 class. All of these students then entered the regular GED preparation program. At the end of the first semester only nine (26%) had dropped out—half the usual rate.

Discussion

In today's political climate expectations for adult education programs are quite high: in a short period of time an educational program should make a measurable impact on the skills or status of adults who — while motivated to improve their lot in life — have a history of failing to meet the requirements of any educational program in an expeditious way. The new National Reporting System for adult education requires that an instructional program demonstrate that a participating student makes a major situational improvement—secure employment, secure a better job, or earn a credential such as a GED. It is also acceptable to demonstrate that a learner advances one "educational functioning level" after taking a class, but the definition of this term is fuzzy at best. These requirements are forcing evaluators of adult education programs to select their measures carefully. Administrators making program adoption decisions now want to know whether adults who might study a new program will make progress on one of these criteria.

Standardized assessments such as the CASAS and TABE have published norms that specify educational levels. But these assessments do not always match the outcomes promoted in new programs. Indeed, tests like the CASAS were developed to measure progress in their own proprietary instructional system, not progress toward widely accepted goals for adult learners.

Compared to K-12, adult educators and policy makers are more interested in performance than in mere knowledge gain. For example, they would rather see that a student could prepare a resume that would impress an employer than pass a test on the principles of good resume writing. But performance—especially in the area of job seeking and workplace behavior—is not easy to assess. It requires trained observers making judgments about adult performance in a job interview or in a workplace setting. This usually leads to a tradeoff for an evaluator between having a very small sample of research subjects that can be carefully observed, and having larger samples but using paper-and-pencil measures.

It is useful to distinguish two types of paper-and-pencil measures: knowledge tests and performance measures. The latter has more credibility in adult education. But performance measures are difficult to construct and score, and they present a higher learning threshold. By definition, a performance task requires both discrete knowledge and the integration of that knowledge to produce a coherent product. Performance tests require more precise learning than knowledge tests, and the learner must practice producing the performance until it meets the criterion.

The limited reading skills of many adults calls for the use of face-to-face interviews in many studies. Limited reading ability affects respondents' ability to follow instructions on forms and to even understand the meaning of a response scale. Needless to say, face-to-face interviewing adds considerably to the cost of collecting data.

Studies of ABE students are often more difficult to conduct than K-12 studies. The research target is not a captive audience coming to school every day. Adults need to be enticed to participate in an educational program and in a study of its worth. If they do participate, they may drop out of the program if their life situation changes. This makes it difficult carry out an evaluation in a tight time frame. The most reasonable design may be a product validation, where a small number of learners are recruited and nurtured to stay with the program. They are given every incentive to engage the materials and stay with the research protocol. Although such a test does not yield "typical" results, it does provide a measure of "possible" results. (See Johnston, 1981, for a discussion of the product validation model).

References

GED Testing Service (2001). The GED Test. Retrieved February 1, 2001 from the World Wide Web: http://www.acenet.edu/calec/ged/

Johnston, J. (1981). Evaluation of curriculum innovations: A product validation approach. In C. B. Aslanian (Ed.), *Improving educational evaluation methods: Impact on Policy* (pp. 79-100). Sage Research Progress Series in Evaluation, Volume 11. Beverly Hills: Sage.

Johnston, J., Young, S.J., & Petty, L., (1999). *An impact study of TV411: Facilitated group design—pilot test of procedures.* Ann Arbor: University of Michigan, Institute for Social Research.

Johnston, J., Isler-Petty, L., & Young, S. J. (2000). *Using TV411 at home: A pilot test of home and outreach models.* Ann Arbor: University of Michigan, Institute for Social Research.

Johnston, J., Isler-Petty, L., & Young, S. J. (2001a). *An impact study of TV411: Using TV411 in a facilitated group.* Ann Arbor: University of Michigan, Institute for Social Research.

Johnston, J., Young, S. J., & Isler-Petty, L. . (2001b). *An impact study of LiteracyLink's Workplace Essential Skills.* Ann Arbor: University of Michigan, Institute for Social Research.

OVAE (2000). *Measures and Methods for the National Reporting System for Adult Education: Implementation Guidelines*. Washington: U. S. Department of Education, Office of Vocational and Adult Education. [web reference]

Star Schools (2001). *Star Schools projects.* Available online: www.ed.gov/EdRes/EdFed/Star.html

U. S. Department of Labor (1992). *Secretary's Commission on Achieving Necessary Skills (SCANS) Report for America 2000.* Washington, DC: Author.

Wigfield, A. (1997). Reading motives: A domain-specific approach to motivation. *Educational Psychologist, 32,* 59-68.

4. Teacher Outcomes: Changed Pedagogy

Marcie J. Bober
San Diego State University

For many years now, the federal government has invested heavily in systemic educational reform. Technology Innovation Challenge Grants (TICG) and Stars Schools – comprehensive programs overseen by the Office of Educational Research and Development (OERI) – are but two vehicles for spurring instructional innovation. Since 1995, for example, TICG funds have enabled school districts, both large and small, to work in partnership with business members, government agencies, and institutions of higher education to change public and private schools in three critical areas: curriculum, professional development, and infrastructure.

Most evaluative efforts have been basic – determining technology's impact by measuring *changes in skill* with selected applications; *equipment distribution* on school campus; or the amount of *time* students and/or teachers spend using computers or accessing the Internet. Students are queried about their *confidence* with technology, their *readiness* for college or the workplace, and whether or not they *like* school.

Many of the evaluators associated with Challenge Grants (and other technology-infusion efforts) are now interested in issues of greater substance. *How … and in what ways … do access to and use of technology alter classroom practices? Is teaching transformed?* Underlying evaluation that focuses on pedagogical change is the belief that technology – when well-used – can lead to performance outcomes that may not be adequately measured by grades, standardized test scores – or, in the case of teachers – traditional personnel assessments: the ability to problem-solve, to make decisions, to adapt quickly to change, to innovate, to work well independently as well as in a team or group setting, to think creatively.

This chapter describes some of the complexities associated with measuring teacher outcomes in terms of pedagogical growth, and it is organized around several major themes:

· What is pedagogy? What does it mean to be pedagogically innovative? What constraints face evaluators tasked with measuring pedagogical change?

- How is pedagogical change related to the technology-focused professional growth activities in which teachers are engaged? What are the characteristics of innovative professional development? How well does professional development attend to competence?

- How is pedagogical growth measured? On what “standardized” tools do schools and districts rely? What project-specific examples are worth emulating?

What is Pedagogy?

By definition, pedagogy is the art, science, or profession of teaching (Merriam-Webster , 2001) On a more practical level, pedagogy refers to the approaches and strategies that guide instruction, as well as the theories that frame them.

Pedagogy: Different Perspectives … Different Outcomes

Before measuring pedagogical change, the first step is to understand the school district’s or project’s pedagogical philosophy. Measurement of pedagogical change must account for the theoretical orientation driving the instruction.

Many educators concerned with technology’s influence on pedagogy tend to embrace *cognitivism*. They are concerned with the mental changes that occur during instruction, as well as learner autonomy and initiative. They ask, *How is information received, organized, retained, and used by the brain?* (Simonson & Thompson, 1997, p. 41). Unlike their *behaviorist* colleagues (concerned mainly with observable indicators of learning and what those observations imply for teaching), cognitivists are focused on the content of instruction itself. Several pedagogical issues arise when the orientation is cognitive, including the learner’s predisposition to learning; his or her developmental stage; the structure and form of knowledge itself; the sequencing of instructional materials and events; the form and pacing of practice, assessment, and reinforcement; and the level/nature of learner control.

Like their cognitivist colleagues, *constructivists* are learner-focused. They argue that each of us has a unique view of how the world works. Authenticity and relevance are but two of the pedagogical implications of a philosophy which contends that people construct meaning based on personal or individual experience. Instruction, constructivists argue, must be content- and stimuli-rich, embrace visual formats, promote teaming and collaboration, and be flexibly organized.

Teacher Certification: Pedagogical Change via Mandate

From a practical standpoint, however, pedagogy is primarily influenced by the undergraduate education a teacher receives – which, of course, reflects state-specific certification requirements (see , University of Kentucky, College of Education, 2001). Today, many states are actively engaged in systemic reform of preservice education – spurred mostly by public perception that teachers are under-prepared to meet their students' instructional needs.

In some cases, reform focuses on *curriculum*. California is but one of several states with well-defined standards-based frameworks for all content areas. In fact, some California districts (San Diego Unified, for example) have enacted grade-level content and performance standards far more precise than the state itself demands. Curricular change in Illinois strives to better align teacher education (Content-Area Standards for Teachers and Administrators) with student expectations (Illinois Learning Standards for Students) (see Illinois State Board of Education, Division of Professional Preparation, 2001)

In some cases, reform focuses on *examination* – in effect, either raising the scores prospective teachers must earn on basic skills and other gateway tests (the route Missouri has taken) or establishing basic competency thresholds (the path Massachusetts has pursued).[1]

But many states, California and Texas among the most prominent, are embracing technology as part of a total restructuring effort. Effective January 2001, California's teachers (both multiple and single subject candidates) must meet specific technology standards at two different points in time: *basic competence* to earn a Level 1 (preliminary) credential, and *advanced competence* to earn a Level 2 (professional clear) credential (see, California Commission on Teacher Credentialing, 2001). While a number of the "competencies" are skewed to the functional end of the spectrum (e.g. each candidate demonstrates competency in the operation and care of computer related hardware), several are clearly pedagogically-oriented (e.g., each candidate designs, adapts, and uses lessons which address the students' needs to develop information literacy and problem-solving skills for lifelong learning).

[1] Although this approach seems fairly logical on its surface, teacher educators in Massachusetts claim an array of unintended consequences – among them, that the public views teachers as second-class professionals unable to pursue "better" careers, that colleges have tightened admissions standards to the extent that underrepresented applicants are systematically excluded/screened out, and/or that colleges are teaching to the test and so eliminating a focus on the "intangibles" [caring, perseverance] that characterize successful teachers (Filippo & Riccards, 2000).

Technology-focused reform underway in Texas means a complete restructuring of existing standards (attending to what educators *should know* and *be able to do* – rather than describing courses to complete), and developing standards for which no previous competencies were articulated (Technology Applications: Grade 4, Grades 4-8, and Grades 8-12). As in Illinois, "each set of [teacher] standards are grounded in the Texas Essential Knowledge and Skills (TEKS), the state's required curriculum for public school students" (Texas Classroom Teacher's Association, 2001). The most recent version of the proposed technology standards targets 11 specific areas; while the final six are specific to "disciplines" (e.g., the web-mastering teacher – Standard XI; the video technology teacher – Standard X), the first five broadly apply to all teachers, for example:

- Standard II: All teachers identify task requirements, apply search strategies, and use current technology to efficiently acquire, analyze, and evaluate a variety of electronic information.

- Standard III: All teachers use task-appropriate tools to synthesize knowledge, create and modify solutions, and evaluate results in a way that supports the work of individuals and groups in problem-solving situations.

While the impact is not immediate (and cannot yet be fully anticipated), revisions in credentialing requirements cannot help but alter the way evaluators assess federally-funded technology projects in place throughout our nation's schools. This author believes the results will be largely positive; as teachers enter the profession with greater functional and applied proficiency, evaluators must, of necessity, refocus their energies on oft-neglected issues that deeply affect pedagogy, e.g., environmental barriers to technology integration (administrative support, resources, time); fear of innovation and change (Rogers, 1995); and accommodating learner differences (borne of unfamiliarity with the language, differences in natural ability, or socioeconomic disparities).

Pedagogy and Technology Deployment

Even the best of instructional intentions can go awry. Following are some of the conditions that doom the most thoughtfully-designed technology-infusion efforts centered on pedagogical growth.

- *Progressive change in instructional methods does not assure that the technology itself is appropriate for the task at hand.* In fact, Thomas and Boysen's (1984, cited in Simonson & Thompson, 1997) taxonomy[2] of

[2] Experiencing is the lowest level; specific applications or programs are used to set the stage for later learning. Informing is the next level; specific programs or applications are deliverers of information/content. At the next tier is reinforcing – featuring programs or applications used following

educational uses for the computer was born of frustration with the (still-common) practice of organizing or classifying technology ‘use’ by software type. A classification system focused on the state of the learner with respect to the material recognizes that the same piece of software may actually serve multiple instructional purposes.

- *Progressive change in instructional methods does not ensure thoughtful technology planning*. In fact, the potential to improve instructional pedagogy is lost when planning for technology use/integration fails to consider distinguishing characteristics of the learning environment into which specific hardware and software will be deployed. Tomei (1997) argues for a focus on the key constructs that form the foundation of instructional planning – specifically, how individuals learn (or "master") content **and** content presentation (including the extent to which it "matches" preferred learning styles and developmental stage). He describes the ways in which instructional technologies support different learning and presentational configurations and how misapplications of IT – instances where deployed technologies ill-fit the instructional setting – result in learner dissatisfaction that ranges from boredom to frustration. Tomei proposes that technology purchasing and implementation be preceded by a thoughtful process in which teachers examine a variety of learning paradigms, determine which of them best suit the instructional settings in which they find themselves, and then select instructional technologies that allow those paradigms to flourish. The alternative, he argues – thinking first (or in isolation) about the technology (and its deployment) – will likely result in selection of tools completely at odds with the instructional premise or framework.

- *Progressive change in instructional methods does not ensure a big-picture vision.* When the pedagogical view is school-centric – focused exclusively on what teachers do for, to, or with students – critical social factors attendant to technology use are unintentionally ignored. Pedagogy may actually be situational, meaning it reflects the needs of real students and the communities in which they live. Guiney (1999) suggests that school policies – be they focused on facilities, human resources, or instructional approaches – mirror the morals, ethics, and values of the constituents they serve. School policies, then, are what result from the interaction (some might say *clash*) of politics, pedagogy, and values. Those who study the business side of education are beginning to see that the pedagogical elements factoring into technology purchases and deployment are heavily

instruction to strengthen specific learning outcomes. At the integrating level, selected programs/applications let students apply previous learning to new situations … and to associate previously unconnected ideas. At the utilizing level, the computer is a tool in the manipulation of subject matter.

influenced (sometimes subconsciously) by politics, the power to control ideas and information, and community image (op. cit., p. 33).

Clearly, then, technology can influence learning – no matter what the frame of theoretical reference. As important, pedagogical change – whether or not specifically measured – is an implied or direct outcome of any technology training in which teachers are engaged. Roblyer, Edwards, and Havriluk (1997) speak of the shifts in behaviors and attitudes that result from consistent exposure to technology and promising techniques for successful classroom infusion/integration. They suggest that through integrating technology into instruction, teachers become more student-centered, less interested in whole-class instruction. The activities and projects they assign are more interdisciplinary and more open-ended – with students encouraged to pursue creative and appropriate solutions – rather than the *right ones*. They stress cooperation and healthy competition.

Professional Development

Changes in instructional pedagogy are directly linked to the professional growth efforts in which teachers are engaged. Thus any evaluation of changes in pedagogy needs to include consideration of professional development strategies and activities.

Professional Development: A Conceptual View

Bellanca (1995) focuses on professional growth conceptually, distinguishing *professional development* from more traditional activities (*staff development or in-service training)* that teachers attend by mandate or choice.

- *Staff development* aims to remedy perceived deficiencies in knowledge or performance. It is a “spray paint” effort to promote innovation, sometimes including demonstrations and opportunities for guided practice. While attendees are encouraged to apply what they have learned, no formal follow-up activities are specifically scheduled, and evidence of changed classroom practices is neither required nor expected.

- *In-service training* is the attempt to promote awareness about new ideas in the field. At its worst, it provides a forum for brief explanations of mandates – curricular or procedural. In-services tend to be brief – often a day or less; the audience captive. The content tends to be general – structured to conform with lecture-style delivery. While some presenters may speak with flair and confidence, it is often left to the individual attendee to determine how the information relates to his or discipline (e.g., science) or student population (4th graders; children with special needs).

- *Professional development*, then, is what allows for constructive educational change and reasoned accountability. It is a planned, comprehensive, and systemic program of goals-driven (often competency-based) activities that promote productive change in individuals and school structures. The focus is on all facets of program life: design, implementation and assessment. Behavioral and attitudinal changes are both expected and supported; although differential involvement is accepted, an array of incentives and rewards promote commitment. Because the effort is systemic, activities are interrelated and cumulative. As important, they complement the school and district's vision/strategic mission and reflect all key constituencies: certificated and credentialed staff, administrators, students, parents, community members.

The views of Sparks and Hirsh (1997) mesh well with Bellanca's. They argue that today's schools – and the professional growth opportunities that occur within them – are driven by three powerful ideas: results-driven education, systems thinking, and constructivism.

- A school or district focused on *results* looks beyond courses offered or grades students receive. What do students actually know, and what can they do as a result of their time in school? Not surprisingly, a results-driven environment means changed thinking about what constitutes successful professional development; indicators related to behavioral or attitudinal change that result in benefits to students far outweigh the amount of seat time or the total number of attendees.

- A school that thinks *systematically* looks at school reform holistically. Reactive thinking that attends to "hot spots" and quick-fixes is replaced by proactive thinking, which promotes interrelationships and interconnectedness among school functions and personnel. Not surprisingly, a school environment that focuses on systems promotes multileveled, well-coordinated professional development that targets everyone – from clerical staff to the principal.

- A school that is *constructivist* recognizes that knowledge is constructed in the mind of the learner, whether that learner be a child or a staff member. The implications of constructivism for professional development, though relatively obvious, are fairly profound. Eclectic classrooms that promote active learning and student autonomy and initiative *are not created* via professional growth activities premised on the transmittal view of learning (lecture, reading silently). A constructivist approach to staff development promotes a collaborative spirit, an action-oriented agenda, and reflective practices.

Professional Development: From Ideas to Practice

To what, then, does strong professional development attend? Danielson (1996) advocates a framework for professional practice that brings clarity to new theoretical paradigms for staff development. Organized into four domains of teaching responsibilities,[3] the framework makes a definitive statement about teaching as a field on par with others we hold in high regard: physicians, accountants, architects. By establishing definitions of expertise and procedures to certify both novice and advanced practitioners, educators guarantee to the larger community that its members "hold themselves and their colleagues to the highest standards" (op. cit., p. 2). Though some might argue with the simplicity of the rating scale (*unsatisfactory*, *basic*, *proficient*, *distinguished*), the structure Danielson advocates attends well to the complexities of teaching as well as its physical and mental demands. Certainly the framework promotes accountability, but more importantly, it "offers the profession a means of communicating about excellence" (op. cit., p. 5) and the different paths its practitioners may take to reach their potential.

By the latter half of the 1990s, both the National Research Council and the National Council of Teachers of Mathematics had released new publications focused on professional development. Though their efforts were not collaborative, their vision and purpose were remarkably similar: to redress well-identified weaknesses in teacher preparation and general teaching practices that stymied students' understanding and appreciation of science and mathematics. Specifically targeted was the over-reliance on lecture to transmit a body of knowledge and skills; the distorted view of each field as little more than a collection of memorizable facts and rules; and an overemphasis on technical skills rather than decision-making, reasoning, and hypothesis-generating. The tenets of professionalism each organization promotes are remarkably discipline-neutral,[4] integrating well into Danielson's four domains *and* lending themselves to the types of growth activities promoted by Bellanca (1995) and Sparks and Hirsh (1997).

Technology-Focused Professional Development: Voices from the Field

According to Norman (1999), top-notch professional growth programs (no matter what their topic or purpose) are always focused on students as the critical stakeholder

---

[3] Domain 1: planning and preparation; Domain 2: classroom environment; Domain 3: instruction; Domain 4: professionalism.

[4] Examples: the ability to use student data, observations of teaching, and interactions with colleagues to reflect on and improve teaching; the ability to implement collaborative and independent tasks that challenge student thinking; the ability to let students have a voice in instructional decisions, specifically about the content and context of their work.

group. *Students are the ultimate beneficiaries of any training teachers receive.* However, other stakeholders play prominent roles in the design, implementation, and assessment of program quality: teachers, administrators, parents, the school board, community members. Planning is strategic, not merely tactical; application-specific skills are far less important than curriculum, instructional strategies and techniques, and assessment. Activities are well-funded, allowing for training customization, ongoing mentoring, and follow-up. There is a focus on modeling that leads to replicable communities of practice. Finally, evaluation is highly valued; program staff recognize that periodic review is positive, not punitive – and integral to sustainability. Continual assessment ensures that the effort reflects changes in school or district policies – accommodating newly-funded initiatives as well as personnel transitions.

Unfortunately, evidence of successful technology-focused professional development is scant and largely anecdotal. Sherwood (1999) describes a mentoring project in rural New York that funded the hiring of a technology teacher who, over a three-year period, worked directly with teachers on tool familiarization, curriculum design, instructional delivery, and alternative assessment. Teachers, in effect, grew with the project, becoming more technologically savvy – and, as important, more independent – as time passed.

Grant (1999) reports on a project conducted abroad by TERC, targeting mathematics education in four Department of Defense elementary schools in Germany. All training was project-based. Teachers learned core application features and functions in the context of solving real problems; at all times they worked with real data. In essence, trainers were as interested in helping teachers become "content competent" as they were with improving their technological savvy. To build confidence, independence, reflectiveness, and collegiality, the trainers opted for a *co-teaching* professional development model premised on action planning and shared strategies for tackling an array of technical and instructional issues. Teachers modeled the very behaviors they wanted their students to embrace.

But more typical are the results of a workshop conducted by the RAND Corporation's Critical Technologies Institute (Harvey & Purnell, 1995). While not ungrateful for the learning opportunities presented to them, participants were clearly negative about the structure and process by which their training unfolded. A common feeling across sites was that technology was divorced from school culture, which allowed it to became the engine that drove curriculum and student needs, rather than the reverse. The energy and impetus – perhaps motivated by visions of "quick-fixes" or the desire to appear cutting-edge to the outside world – were top-down rather than bottom-up. Little effort was made to invest teachers in the process or the outcomes, or to grow an infrastructure that valued networking and information-sharing. As critical was that the focus was incremental and small scale; competence – central to any meaningful growth in instructional pedagogy – was never the goal.

And perhaps it is the failure to strive for *competence* that best explains why so many well-intentioned professional growth efforts go awry. Perry (1998) describes competence as a cluster of related *knowledge, skills*, and *attitudes* that affects a major part of one's job (i.e., one or more key roles or responsibilities); correlates with performance on the job; can be measured against well-accepted standards; and can be improved via training and development. It should come as no surprise, then, that performance gains (manifested in the present case by innovative, learner-centered classroom practices) are severely compromised when staff development fails to attend to all three elements of the cluster, or participants are not aware that their training has been of limited focus.

Competence, Perry argues, is generic and applies in many situations; there is more opportunity to apply generic competence than any specific skill. Competence is what allows an educator to move easily from teaching 3rd grade one year to 5th grade the next … from using PC-based computers to Macintoshes … from 50-minute periods to block periods. It is competence that ensures a teacher's ability to deal with software glitches and the constant barrage of new features and functions. Readiness and willingness to adapt to change are at the heart of technological innovation; they are also at the heart of pedagogical excellence.

Aiming for pedagogical competence means placing less emphasis on technical savvy (prowess with specific software applications, for example) and more emphasis on audience analysis, adaptivity, and flexibility (i.e., responsiveness to students' individual needs). A teacher whose technology training has positively influenced pedagogy knows how to manage instructional time, how to organize students to take advantage of technological resources, how to ensure equitable access, how to encourage student initiative. Thinking about professional development as constructivist in nature and competency-based can dramatically alter *how we plan for* and *what we expect from* technology training.

But even stories of success – where participants leave a training experience motivated to alter old instructional habits – can be clouded by the growing emphasis on standards-based instruction and standardized tests as valid measures of pedagogical excellence. Herr (2000), for example, speaks glowingly of the ways in which managed technology solutions (specifically, CCC's SuccessMaker) allowed her students to excel on Virginia's Standards of Learning tests – the results of which determine whether or not a student matriculates to high school and, once there, earns either a diploma or merely a certificate of attendance. With such pressure to perform – with so much on the line – Herr focused her attention on aligning her curriculum to the tests upon which her students' very futures depend. Her innovativeness, though commendable, speaks far more of her ability to master and manipulate the nuances of the SuccessMaker management system than pedagogical growth that fosters student autonomy and self-directedness.

The fallacy of using student performance to measure instructional pedagogy is evidenced via a case study reported by Goldman, Cole, and Syer (1999). As part of the **Challenge 2000 Multimedia Project**, teachers used the tenets of problem-based learning (PBL) to develop multimedia-enhanced units of instruction that foster subject-matter knowledge, technology savvy, and collaboration among peers. While teachers themselves were provided an array of professional growth opportunities – helping them become more conversant in specific technologies as well as core components of PBL – students themselves had no direct technology instruction. As a result, rich instructional opportunities were lost as students struggled with immature search strategies while seeking content or fell into labor-intensive methods for making changes to their work (for example, rekeying entire blocks of text rather than turning either to the software's basic editing functions or advanced revision features).

"Standardized" Tools for Measuring Pedagogical Growth

There is clear evidence that pedagogical excellence is increasingly important to grant funders – the federal government among them – even if current systems for measuring it are conceptually weak, simply inappropriate, or mired in political whimsy.

NETS

The *National Educational Technology Standards* (*NETS*) for teachers advocated by the International Society for Technology in Education (ISTE) are touted by an array of government agencies and professional associations – and serve as the benchmark by which a growing number of federally-funded professional development efforts are assessed (See ISTE, 2001). Three of the six *NETS* clusters attend to issues pedagogical in nature.[5]

[5] II – Planning and Designing Learning Environments and Experiences (Teachers plan and design effective learning environments and experiences supported by technology); III – Teaching, Learning, and the Curriculum (Teachers implement curriculum plans that include methods and strategies for applying technology to maximize student learning); IV – Assessment and Evaluation (Teachers apply technology to facilitate a variety of effective assessment and evaluation strategies)

At this time, however, *NETS* are largely geared toward teacher preparation. Thus the proficiency scale reflects where pre-service teachers may be as they proceed through their methods courses, field experiences, and first year of professional practice. The implication is that teachers whose *entire preservice education has been technology-infused* are as well prepared to deal with technology issues as they are with those far more typical of teacher education (instructional planning, content presentation, assessment). But it is important to note that *NETS* for teachers do not presently include strategies for measuring competence or attainment of standards *or* advocate training interventions that promote competence or proficiency. The path to excellence, whether pre-service or in-service, is left to individual schools, districts, and states to fashion.

CTAP

The *California Technology Assistance Program* (see, CTAP, 2001) is tasked with providing assistance to schools and districts integrating technology into teaching and learning. It promotes effective use of technology through regional coordination of support services based on local needs organized around five core areas: staff development, technical assistance, information and learning resources, telecommunications infrastructure, and funding. Like *NETS*, *CTAP* also promotes leveled sets of proficiencies[6] – but the approach is far more prescriptive. Each general knowledge or skill area (e.g., Communication and Collaboration; Planning, Designing, and Implementing Learning Experiences) is organized around performance indicators that, in some California regions, have resulted in check-off validations of competence.

UTAP

The *Utah Technology Awareness Project* (*UTAP*) uses an online self-assessment tool for educators – in essence, a collection of rubrics tied to six specific skill areas: basic concepts, personal/professional productivity, communication/information, classroom instruction, educational leadership, and administrative leadership. Informed by research conducted under the ACOT umbrella, *UTAP* impacts both in-service and pre-service teacher training. “It helps teachers see what they need to learn to effectively leverage technology in the classroom” (see, UTAP, 2001).

Once registered in the network, teachers complete one or more of the surveys, view results, and then select professional develop opportunities “customized” to

[6] See, for example, the Preliminary Technology Proficiency file, at http://www.fcoe.k12.ca.us/techprof/preliminary_profiles1.htm, the Professional Technology Proficiency file, at http://www.fcoe.k12.ca.us/techprof/professional_profiles.htm, and the Technology Certification Proficiency Checklist promoted by Kern County, at http://www.ctap.org/ctc/download.htm).

identified weaknesses and strengths. Ongoing skill assessment is encouraged; higher scores on the surveys imply the effectiveness of the activities they have opted to attend/complete. The Measurement Appendix includes screenshots from this tool that depict the assessment path a typical teacher might follow to determine his or her proficiency relative to selecting training classes that build classroom instruction skills.

Flashlight

The TLT Consulting Group helps institutions develop realistic and measurable technology goals. Its best-known product is the *Flashlight Program*, a proprietary inventory of approximately 500 survey items that licensed users can tap to create customized questionnaires aligned with specific evaluative purposes. The survey items are organized into several strands, among them:

- *Educational Strategies*, which focus on teaching and learning practices (for example, how selected practices [e.g., collaboration] are helped or hindered by specific technologies [e.g., audioconferencing or email]).
- *Experiences with Technology*, which focus on technology use (e.g., how often, within a specified time frame, a student uses various technologies *or* for what purposes or reasons different technologies are used) and technological sophistication (e.g., proficiency with specific applications or programs).

The interview/focus group protocol is organized around such issues as *active learning*, *collaborative learning*, *time on task*, *respect for diversity*, and *cognitive and creative outcomes*. While Flashlight participants tend to be students, clients often use the results to plan their professional development programs.

Seven Dimensions

The Milken Family Foundation for Educational Technology promotes its *Seven Dimensions for Gauging Progress,* helping policymakers, educators, and technology directors determine "… the conditions that should be in place for technology to be used to its greatest educational advantage in any classroom" (Lemke and Coughlin, 2001, p. 158). A companion document (*Technology in American Schools: Seven Dimensions of Progress – An Educator's Guide*) includes a continuum of progress indicators for each dimension, organized around three stages of progress: entry, adaptation, and transformation. Transition steps guide the educator from one stage to the next.

Each of the framework's independent dimensions features fundamental questions that stakeholders within a K-12 environment should consider as technology and telecommunications are deployed.

- *Dimension 1* – Learners: Are learners using technology in ways that deepen their understanding of the academic content and advance their knowledge of the world around them?
- *Dimension 2* – Learning Environments: Is the learning environment designed to achieve high academic performance by students (through the alignment of standards, research-proven learning practices and contemporary technology)?
- *Dimension 3* – Professional Competency: Are educators fluent with technology and do they effectively use technology to the learning advantage of their students?
- *Dimension 4* – System Capacity: Is the education system reengineering itself to systematically meet the needs of learners in this knowledge-based, global society?
- *Dimension 5* – Community Connections: Is the school-community relationship one of trust and respect, and is this translating into mutually beneficial, sustainable partnerships in the area of learning technology?
- *Dimension 6* – Technology Capacity: Are there adequate technology, networks, electronic resources, and support to meet the education system's learning goals?
- *Dimension 7* – Accountability: Is there agreement on what success with technology looks like? Are there measures in place to track progress and report results?

For purposes of this chapter, the focus is on Dimension 3: Professional Competency. Here, the core or fundamental question (posed above) centers around four areas or strands: (1) core technology fluency; (2) curriculum, learning, and assessment; (3) professional practice and collegiality; and (4) classroom and instructional management.

The *Professional Competency Continuum Online Assessment Tool* – helps educators (teachers and administrators) assess their status within the skill and knowledge areas showcased in the Professional Competency Dimension. The General Assessment provides an overview while Detailed Assessments in the four major areas or strands generate customized advice and resources. Those responsible for professional development are encouraged to create a "project" – allowing a variety of group and individual reports to be generated. A portion of the self- assessment featured within the curriculum, learning and assessment strand can be found in the Measurement Appendix.

STaR Chart

Finally, the CEO Forum on Education and Technology, founded in 1996 with an eye to helping educators better prepare students to be contributing citizens and productive employees (see, CEO Forum, 2001) offers its *School Technology and Readiness* (*STaR*) *Chart*,[7] a tool for benchmarking/monitoring progress toward technological readiness and proficiency. The framework is built on three key questions: *Is technology (at the school or district level) being used in ways that ensure the best possible teaching and learning? What is a school or district's "technology profile?" What areas (at the site/district level) should be targeted to ensure effective integration?*

The *STaR Chart* produces four school profiles that range from *Early Tech* (a school or district with little or no technology) to *Target Tech* (a school or district that serves as an innovation model for others to emulate). Schools report using the *STaR* results in several ways: to set goals (and then monitor progress toward their attainment); to identify technology needs for which grant/award applications may be written; to determine how best to allocate technology funds already available; as the basis of statewide technology assessments. Only a very few questions are pedagogically-oriented. Nonetheless, the survey (which is simple to deploy and offers results and next-steps prescriptions that are easy to interpret, is increasingly used as a needs assessment tool that spurs professional development planning. Items focusing on pedagogy include the following:

What are the capabilities of teachers?

- Basic technical skills including applications such as word processing/Little or no use in instruction
- Utilize standalone software/Employ some Internet and e-mail
- Integrate digital content into instruction /Manage classroom learning with technology
- Create a digital learning environment

How do teachers integrate digital content to enhance instruction?

- Use as a supplement rather than an integral part of the curriculum
- Use to streamline administrative functions, communicate, and for presentation
- Use for research, lesson planning, multimedia and graphical presentations and simulations, and to correspond with experts, peers and parents
- Digital content changes the teaching process, allowing for greater levels of inquiry, analysis, interest, collaboration, creativity and content production

[7] An adaptation of the *STaR Chart* targets teacher education programs. In terms of purpose/intent and output (i.e., four evaluative profiles), it is identical to the K-12 version.

What is the role of the teacher?

- Teacher-centered
- Teacher-directed
- Teacher facilitated
- Teacher as guide; Student-centered

What is the frequency of instruction using digital content?

- Weekly
- 3-4 times a week
- Daily, but activities are isolated by grade, disciplines, classes
- Seamlessly integrated throughout all classes and subjects on a daily basis

Examples from Star Schools and Challenge Grant Evaluations

In an earlier chapter on measuring cognitive outcomes for students, Cassidy points out several techniques for direct (e.g., standardized tests, tailored tests, alternative assessments, or performance tasks) and indirect (e.g., reports of parents and teachers, self-reports, and archived data) measurement of student outcomes in the cognitive domain. Each, we learn, has strengths and weaknesses: each is more or less time-consuming to conduct; each is more or less accurate in estimating actual ability (knowledge or skill). Measurement of technology's impact on instructional pedagogy is similarly problematic, although a number of Challenge and Stars Schools evaluators – especially those evaluating mature projects – are attempting to do so. Among the tactics they employ are surveys, case studies (which feature observations, interviews and focus groups), and artifact collection. The following examples – drawn from the Challenge Grant program – provide some insight into instrument design/structure, implementation or administration, and data value.

Surveys

ACT Now! (Sweetwater Union High School District)

The **Advanced Curriculum through Technology program (*ACT Now!*)** is a district-wide effort that focuses on professional development and infrastructure build-out. Teachers who agree to attend 40 hours of training over a six-month period during the school year (24 units of core courses; 16 units of electives) receive a high-end multimedia computer, an array of software, and site-based assistance (both technical and curricular).

Workshops are largely conducted by a hand-picked team of teachers – each of them recognized in the field for his or her excellent teaching skills, technology savvy, content or disciplinary expertise, and leadership ability.

In 2000 (and again in 2001), a revamped end-of-school-year survey was administered to teachers who had met their professional development commitment between August 1998 and May 1999, August 1999 and May 2000 *or* August 2000 and May 2001. Survey revisions were designed to focus specifically on project objectives that speak directly to transformed teaching. Most questions were placed within one specific section of the survey, and featured the following ordinal scale:

- Isn't really part of my everyday practice
- Is generally a part of my everyday practice
- Is fundamentally a part of my everyday practice
- Is integral to my everyday teaching practice

The items were designed to reflect specific characteristics of constructivist teachers described in the ASCD publication, *The Case for Constructivist Classrooms* (Brooks & Brooks, 1999), the 22 components of professional practice described in Danielson's (1996) framework, and specific professional competencies advocated by the National Research Council (1996), and the National Council of Teachers of Mathematics (1991). The survey also featured several items that tangentially attend to ASCD guidelines or tenets of practice advocated by respected professional organizations but whose relationship to *ACT Now!* goals and objectives – and the strategies/techniques demonstrated in nearly every workshop teachers attended – was clear. Below are eight of the 30 items in the survey. (The full survey can be found in the Measurement Appendix.)

- Allowing my students to contribute to the decisions I make about the content and context of their work
- Helping to plan and implement professional growth opportunities for teachers at my site
- Allowing my students to contribute to the decisions I make about the content and context of their work
- Implementing lessons and units that are standards-based
- Allowing my students to contribute to the decisions I make about the content and context of their work
- Modeling the skills of inquiry – including skepticism, curiosity, an openness to new ideas, and an interest in data
- Assessing both student understanding and student skills

- Using multimedia technologies to create materials that students use in class

Admittedly, surveys are an indirect measure of pedagogical growth. Nonetheless, the data provide perceptual evidence of strengths and weaknesses – and cannot be dismissed out of hand. An analysis completed in December 2000 suggested that teachers were at the early stages of instructional transformation, and so may not have seen technology's potential to engage students in learning, encourage information sharing, and foster guided independence. While interested in transforming the learning environment, teachers were not yet prepared to actively involve students in decisions related to instructional delivery and assessment. Their focus was on tasks in which students were engaged rather than on the environment in which learning took place. As important, many did not yet see themselves as instructional leaders – either at the building or the district level. Perhaps because they were not yet ready for leadership roles, they did not take advantage of parental influence, and did not encourage parents to play a greater role in the design of effective instructional programs. Finally, teachers did not yet fully understand the broader implications of technology infusion – its ability, when part of a thoughtful pedagogical stance – to level the playing field for all students and foster an instructional environment that attends both to what students learn and how they learn it.

El Paso Challenge Grant

The El Paso Project takes a holistic view of educational reform, using technology as a vehicle to transform the teaching, leadership, *and* advocacy skills of veteran and pre-service teachers as well as school administrators. Training also targets parents, who take the lead in operating and then managing special Centers in participating schools.

In Spring 2000, evaluators with the **El Paso Collaborative for Academic Excellence** administered a survey to spur teacher thinking about ways multiple initiatives work together to make schools a better place for teaching and learning. While many of the questions were similar to those posed to Sweetwater participants; the evaluators chose very different response options.

For the following sample of items, the set was time-based: *Every week*, *Every month*, *5 to 8 times a year*, *1 to 4 times a year*, *never*:

- My students use a variety of technological resources (e.g., Internet, databases, computer networks, video) to gather and synthesize information and to create and communicate knowledge
- My students practice computational mathematical skills in the context of real world problems

- My students work in small groups to actively construct new knowledge based on prior learning
- I hold individual conversations with each of my students where they explain to me, and we discuss, their writing processes or strategies for solving math problems.
- Encouraging the use of computers, calculators, and other technologies
- Using multimedia technologies to create materials that students use in class
- Allowing my students to contribute to the decisions I make about the content and context of their work
- Incorporating multimedia technologies into my teaching
- Modeling the ways technological tools can help students reason, make connections, and solve problems
- Helping to plan and implement professional growth opportunities for teachers at my site

For the following sample of items, the set was a standard Likert scale: *Strongly agree*, *Agree*, *Not sure*, *Disagree*, *Strongly disagree*:

- Mathematics problem-solving instruction should be balanced with skills and concept building.
- At our school, teachers value their own learning by participating in continuous personal development (e.g., seek ways to improve their professional practice through colleagues, seminars, professional reading, professional organizations, and so forth).
- At our school, teachers engaged in constructive dialogue to improve teaching and learning.

Some items employed a preparedness/readiness to use scale (*No experience yet*; *Exploratory stage*; *Use, but at slow stage*; *Skillful, but not perfect*), with a few questions functionally oriented and others (such as the ones listed here) more pedagogically focused.

- Mathematics problem-solving instruction should be balanced with skills and concept building.
- At our school, teachers engaged in constructive dialogue to improve teaching and learning.
- At our school, teachers value their own learning by participating in continuous personal development (e.g., seek ways to improve their

professional practice through colleagues, seminars, professional reading, professional organizations, and so forth).

Generation www.Y

Generation www.Y promotes students as collaborators in the curriculum development process. Middle and high schools who become members of the collaborative offer students a special technology course that fosters development of a broad range of critical skills – information and technological literacy; research; writing; presentation; mentoring; project development; and leadership. Students then partner with teachers to develop instructional units suitable for classroom implementation. Consultants with content and technical expertise provide objective feedback to the teacher/student teams on a variety of issues, among them, project feasibility, the project's educational value; and the alignment of objectives, procedures, and assessment.

In Spring 2000, evaluators with Generation www.Y (and related projects) deployed surveys addressing several domains related to instructional pedagogy: attitude toward student-centered, project-based, authentic, constructivist practices; attitude toward collaboration and team-based work; personal self-efficacy for using networked computers in teaching, learning, and instructional design; classroom use of project-based learning; classroom use of constructivist methods; personal collaboration on curriculum projects; and interactions with teachers (on-site or at different schools).[8]

The September 1999 report indicates that the response sets framing many of the questions were fairly traditional:

- Has the frequency of the following (use of specific applications; use of technology for specific purposes; use of technology for instructional planning) changed as a result of your involvement with the project? (Scale: More frequently, Same frequency, Less frequently)
- How has your comfort level with the following (using computers, helping students use computers, using specific application types) changed as a result of your involvement with the project? (Scale: More comfortable, Same level of comfort, Less comfortable)
- Rate your opinions regarding use of technology in education (for example, to facilitate positive change in classroom teaching and learning practices). (Scale: Strongly agree, Agree, Not sure/Neutral, Disagree, Strongly disagree)

[8] Compilations of survey data may be viewed online at http://www.nwrel.org/eval/genwhy/data/index.html; evaluation reports are available for download at http://www.genyes.org/genwwwy/evaluations/index.php.

Case Studies (Observations, Interviews, Focus Groups)

A number of project evaluators conduct case studies – which allow for more direct measurement of pedagogical change.

Virtual High School Challenge Grant

Virtual High School is a collaborative of high schools from around the country. In exchange for contributing a limited amount of teaching time, a member school can offer its students a broad range of *NetCourses* – some Advanced Placement, some technical, some specialized. An on-site VHS coordinator provides project management and teacher support. Teacher participants commit to attend professional development that focuses on best-practices associated with technology-infused instruction.

Evaluators associated with the project are taking a longitudinal view (four years) of five institutions, distributed across regions and types of schools. Two-day site visits allow for semi-structured teacher interviews and lengthy lab or classroom visits that draw on a guide [not available for review] focused on the learning environment, instructional practices (content, pace, assignments, teaching "style") and student-teacher and student-student interactions. Several additional hours are spent reviewing course websites – including the quality of materials available to students and the ease with which they may be accessed. Though the evaluation reports are brief (~two pages of text to showcase each case), it is clear that evaluators are attending to student characteristics, course context, instructional presentation/pacing, scheduling, use of technology (types of software and hardware; the purpose of which each is employed), and community-building.

Results from the most recently submitted evaluation report (covering the 1999-2000 school year) reveal the power of case studies to address pedagogical impact. Team members found profound similarities between VHS and traditional versions of courses, including similar goals, content, assignments, and rollout. Teachers in both settings were high-quality – strong in their areas of expertise, flexible, and motivated. But there were important differences in the two course versions with consequences for long-term student learning. Time for virtual students was more difficult to manage, meaning students had less time to complete both individual and team-based assignments and projects. There was less interaction in VHS courses – both student/student and student/teacher. Some student products (especially those assigned as part of a graphics class) did not lend themselves to the type of feedback the teacher was able to provide. An immature technology infrastructure – and software with clear functional limitations – constrained collaboration among teacher participants and – more importantly – made excessive demands on teacher time. As a result, pedagogical issues (instructional strategies and methods) were less attended to.

Teacher Led Technology Challenge

The **Teacher Led Technology Challenge (TLTC)** works to provide all classroom teachers (pre-K thru 8th grade) in Berkeley (California) with the tools, support, and professional development they need to make computers and other technologies viable tools for teaching and learning in their classrooms. Through a menu of professional development options, classroom teachers learn how to use technology to accommodate diverse learning styles; promote student collaboration (including cross-age tutoring); promote learner independence and initiative, and involve parents as active learning partners. The scaled approach to training fosters teacher confidence in their ability to use technology effectively – which for some means functional savvy, but for others means personal productivity or curricular integration.

The TLTC evaluation team uses observations to measure technology integration – specifically, documenting technology use in classrooms and its influence on the nature and structure of activities in which students are engaged. The protocol for conducting site visits is multifaceted (see Berkeley Planning Associates, 1997; reproduced in the Measurement Appendix).

Guidelines for the observer cover classroom selection and appropriate observer demeanor, areas on which to focus: hardware availability, student products, student organization, interactions between/among students and between students and the teacher. As important, however, is that the guidelines serve as an observation job-aid, featuring note-taking tips (use of descriptive language, nonjudgmental tone, targeted skills, etc.) and general classroom conduct (unobtrusive to the extent possible, participation by invitation only).

Summary Forms focus on lesson/activity logistics (including goals and objectives, outcomes, and sequencing of events), the scope/context of the lesson/activity itself (length of time to complete, student groups, level/type of student input, tasks that students undertake), teacher/student roles (with a focus on how instruction is facilitated), nature of student behavior, learner accommodations (whether attending to ability/skills, physical needs, or issues of diversity/equity), and technical issues.

Observations (supplemented with interviews) unfold over a period of years. Case study teachers (including several at each grade level across different schools) participate in one or two interviews and observations per year. These observations are supplemented by one-time observations of individual teachers of interest as the evaluation highlights specific intervention strategies or aspects of project design each year. Observers compare notes across interviews and discuss observations emerging from year to year to provide an overall assessment of the nature and quality of technology integration – determined by reflecting back on the project's core goals. For case study teachers, observers update the case studies each year to reflect each teacher's experiences and development over time, developing a better understanding of teachers' developmental process and how it is affected by school environment, project participation and other factors.

The results of these observations show that the effective integration of technology into the regular classroom curriculum is a developmental process that takes most teachers several years to accomplish affectively, but that even in their first year of exposure to the project, most teachers can use technology to support core curriculum in at least a few activities or lessons. The extent to which teachers integrate technology into teaching and learning seamlessly and on a regular basis seems to depend to a large degree on the teaching and classroom management styles of the teachers prior to participation in the project. That is, those teachers who regularly use small group activities and constructivist approaches more readily integrate technology in ways that support cooperative and constructivist learning. Teachers with more teacher-centered pedagogical styles may tend to use technology as a supplemental resource for students to use on their own, or simply take longer to become comfortable using technology creatively as a teaching and learning tool for the whole classroom.

Louisiana Challenge

The **Louisiana Challenge** is part of statewide systemic reform that targets all facets of public education – teacher preparation; curriculum design and implementation; and assessment (Louisiana Department of Education, 1997). Advanced technologies ensure that that state's traditionally underserved students can compete in the educational marketplace – that they are ready to learn; have parents or guardians committed to their academic success; see learning as a lifelong commitment; and have access to the same high-quality instructional materials and top-notch teachers as their more economically-advantaged counterparts. The overall strategy is to build replicable communities of practice.

Case studies were conducted beginning in 1998. Inclusion criteria were fairly specific to ensure data reliability – and meaningful comparisons between project and non-project classes. Day-long site visits (fall and spring) allowed observers to map the classroom environment – including access to various technologies, furniture arrangement (tables, carrels, desks), teacher placement, general room décor (bulletin boards, visual displays), etc. Principals agreed to participate in wide-ranging interviews conducted at specific milestone dates during the school year, and to provide supporting district- and school-level data – including results of standardized tests, attendance and disciplinary information, and descriptions of other initiatives that might confound data analysis and interpretation. Teachers also committed to interviews, as well as artifact collection (student products and portfolios), participation in online collaboratives, and timely submission of surveys and other data elements (including an electronic journal.

In terms of pedagogical impact, the resulting teacher-oriented vignettes tell a compelling story of emerging technology savvy, confidence in using technology effectively, and growing awareness of how technology can lead to student-centered curriculum that meets state mandates but also builds the decision-making and problem-solving skills that characterize the 21st century worker/employee.

Triton

Curriculum development is a primary goal of the **Triton Project**. Teachers participate in scaled professional development that focuses as much on pedagogy as technology savvy. In terms of outcomes, teachers develop, implement and disseminate standards-based instructional units that are relevant to students, cross traditional disciplinary lines, and feature web resources and graphical organizing tools.

Site visits have been integral to project evaluation since 1996. Formal or protocol-driven visits (see the Measurement Appendix), each lasting about 2-1.2 hours, are conducted on a yearly basis by two members of the evaluation team; semi-structured visits are conducted monthly by the project's resource teachers. The evaluation team's intent has been to discuss project progress with key staff at the school. In general, the team a) conducts interviews with the appropriate school administrator; the Site Coordinator; and teachers directly involved in developing and implementing technology-enriched curriculum, and b) tours the classrooms of project teachers – speaking informally with students about their use of technology and noting the types of hardware and software available and overall technology distribution. To prepare for a round of visits, team members skim the school's original project application, the most recent version of its Site Technology Plan, the most recent version of its Annual Action Plan (a district document that details, among other things, the school's learning goals), a list of recent technology purchases (made with project or matching funds), and web-based units of instruction under development or already implemented with students.

Categorically, the site visit protocol attends to project impact (on teaching, the classroom, students, the school, relationships with parents and the larger community); the curricular design process and how it unfolds; successes and missteps with unit implementation; environmental factors (competing mandates, limited administrative support, poor infrastructure) that limit technology use/integration; and alignment of project goals with existing site- or district-level initiatives. All data are entered online – becoming part of a comprehensive network of relational databases that allow data of different types and collected at different points to be search, sorted, and compared.

The evaluation team compiles brief reports of each visit—available for review by the Project Director and project resource teachers upon request. Informally, the team provides the Project Director with an overview that describes successful sites[9] and distinguishes them from those struggling with curricular or project management issues.

Artifact Collection

It is increasingly common to see pedagogical change measured by indicators that clearly are *student associated*, either "performance-based" (i.e., scores on standardized tests) or attitudinal: self-ratings on surveys that assess student confidence with technology; student perceptions of technology's role in preparing them for "the future"; student attitudes toward school and the general learning environment; and the extent to which students report "using" technology to complete an array of tasks. While the information is certainly revealing, it is decidedly secondhand and indirect – and largely discounts critical factors (environment, for example) over which teachers have little control.

Triton

Teacher products – and their quality and reflection of core project objectives – may be a more valid way to approximate changes in pedagogy. A major curricular outcome of the **Triton Project** is the design and implementation of WebQuests (see, San Diego State University, Educational Technology Department, 2000). WebQuests are generally assessed by the project's three technology resource teachers via a multidimensional rubric developed by curriculum advisor Bernie Dodge of the Dept. of Educational Technology, San Diego State University. Available online, the rubric features a weighted scale organized around five specific dimensions (Dodge, 2000):

- Overall aesthetics: generally, the lesson's visual appeal (use of graphic elements to support the content, e.g., concepts, ideas, and relationships;

[9] Characterized by strong administrative support, ongoing/regularly scheduled professional development directly focused on technology use/infusion, consistent technology access (by teachers and students), a collegial spirit among staff, educator confidence relative to technology use, and parental awareness of the school's technology efforts.

variation in layout and typography; color sensitivity; background/foreground contrast; etc.

- Introduction: generally, how well it draws the learner into the lesson and builds on the learner's prior knowledge – effectively preparing him/her for instruction.
- Task: generally, how well the task connects to standards (specifically, what the learner must know or be able to do to attain proficiency) *and* challenges the learner to synthesize multiple sources of information or take an informed position.
- Process: generally, the clarity of the directions the learner must follow; the sufficiency of the instructional scaffolding; and the variety of ways in which a learner must work (independently, with others) to gain perspective and understanding.
- Resources: generally, their usefulness/relevance; their uniqueness/variety; and their time sensitivity.
- Evaluation: generally, the clarity of the criteria and their reasonableness.

Using relational databases helps the technology resource teachers more easily organize the data, and determine quality and appropriateness against such criteria as targeted grade level(s), targeted subject area(s), targeted district or state content/performance standards, special features (e.g., whether or not the unit requires field work or incorporates project partners; the extent to which the unit promotes technology literacy standards or embraces constructivist principles), and repository of student work emerging from the unit. In some cases, student projects have been digitally archived, providing yet another quality-assurance measure.

ACT Now! (Anderson, Indiana)

The **Anderson Community Technology (ACT Now!) Project –** a consortium of local and state agencies in Indiana that "combine" technology initiatives to raise academic achievement – approaches technology integration in multiple ways – through increased parental involvement, the design of new instructional settings, extended day activities (some of them home-based), electronic field trips to local-area cultural areas (including Indianapolis Zoo and Children's Museum), and cable-delivered programming adapted for students with special needs. A comprehensive professional development effort focuses on unit/lesson plan design and development – much of it discipline-focused (social science, language arts, mathematics, and science).

Teachers participating in the project submit lesson plans to program staff at regular intervals – starting at one per month, then one per semester, and finally one per year. Plans are then coded for review by the evaluation staff.

- Coding attends to lesson or unit structure; raters look for mention of particular technological tools; whether or not a specific level of skill, knowledge, or ability is required for successful use of a tool; the primary tool user (teacher, student, both); and whether or not technology is essential to the lesson's instructional integrity (*not a major part*, *somewhat a part*, *integral*).

- Ratings for each coded variable are complex – the idea being to encourage teachers to state their intentions specifically rather than infer them. For example, a lesson for which a tool has *no* prerequisites should include instructions for its use *or* a description of how it is to be introduced and modeled.

A growing database allows project staff to cross-reference lessons in multiple ways; by the project's end a repository of discipline-specific materials – demonstrating ways students and teachers may use a variety of tools to accomplish a variety of tasks – will be available statewide.

Discussion

The measurement/assessment of technology's impact on instructional pedagogy is fraught with difficulty. Recognizing and acknowledging the challenges is a first step in a comprehensive effort to improve the quality of the evaluations of professional development efforts.

Challenges that Derive from the Original Proposals

In some cases, evaluation of impacts on pedagogy is constrained by the original proposals themselves. By today's standards, the very program goals and outcomes that convinced funders just a few years earlier are woefully outdated – overly focused on isolated skills and proficiencies. Nonetheless, there is a tendency for evaluators to continue/follow through with designs that measure growth or change in these areas (e.g., the *ability to identify and select* hardware features and software functions; the *amount* or *proportion* of time students spend using technology). However, the zeal to be faithful to evaluative tasks comes with a price, the most serious of which is missed opportunities to critically examine the real (and often subtle) innovations that spell real project success.

To exemplify, evaluation reports often neglect innovations associated with two key processes: planning and implementation. This stance means that the short- and long-term decisions that characterize these phases are often inadequately described. Few evaluation narratives of which this author is aware speak fully to the ways in which decision-making empowers or disenfranchises key groups (teachers, classified staff, administrators, students, parents) or alters the relationships within and among them. Few evaluation narratives describe new tools designed to help teachers better visualize technology's potential (by letting them examine a variety of learning paradigms, determine which of them best suit the instructional settings in which they find themselves, and select instructional technologies that best allow them to flourish). Better reporting would help school staff see how misapplications of technology – instances where deployed technologies poorly fit the instructional setting – result in learner dissatisfaction that ranges from boredom to frustration. Misplaced concern with original project goals and objectives may mean that we fail to consider in our evaluations the alignment of technology, instructional intent, and pedagogical approach. That omission inadvertently allows the same planning and implementation mistakes to be made again and again.

Challenges that Derive from Terminology

Today's evaluators find themselves at a vocabulary crossroads, which means losing sight of the pedagogical changes most technology projects/programs are designed to evoke. In fact, definitional differences result in evaluations that are perceived to lack focus, rigor, and usefulness.

Several terms – among them pedagogy and competence – were addressed earlier in this document. But the word *technology* itself can be defined in many ways – with each definition bearing its own evaluative connotations. Evaluators constantly struggle with the reference point. Is it hardware? Is it software? Traditionally, technology has been defined as systematic and systemic thinking about the application of knowledge, tools, and skills to solve practical problems, create new products that satisfy human needs, and enhance human performance. But that unique (and perhaps arcane) frame of reference can result in instruments that are not easily normed. If we can't fully agree on a definition of technology, what about the terms *technology integration* or *technology infusion*? While Yepes-Baraya tackles the complexities of measuring integration or infusion (from what it "looks like" to its pedagogical outcomes) elsewhere in this book, the evaluative implication of vocabulary confusion is clear. We simply find ourselves unable to determine instrument reliability and validity in ways that good science demands, and worry that the conclusions we draw are largely unsupportable in the long run.

Challenges that Derive from Confusion over Orientation

Finally, are evaluators thinking strategically when assessing the pedagogical implications (strengths, weaknesses, and impacts) of their projects, or is the view tactical only? Once again, Merriam-Webster is relatively useful for comparing the two terms: tactic: a device for accomplishing an end; strategy: a careful plan or method – devised to achieve a goal.

Planning that is tactical in nature is carried out (more or less) with only a limited or immediate end in view. On the other hand, strategic planning is conducted within the context of an integrated whole; the view is long-term.

By understanding the implications of tactical and strategic planning, evaluators are less apt to examine technology and its pedagogical impact in isolation – rather than as part of comprehensive (multifaceted) school reform. Rather than focus on the obvious and expected (e.g., changes in student performance, unit/lesson facilitation, or assessment), they can attend to issues of substance and with broader implications. How has school culture changed? Are schools safer? Are administrators and other staff more aware of ways to assist students in need? Are parents meaningfully engaged? Is the business community invested? Are faculty more collegial, more involved in building communities of practice?

The bottom line is simply this: while being dutiful to their tasks, many evaluators are tending only to short-term outcomes and effects … rather than seeing the interventions they are tracking as part of a total package designed to change how schools function.

References

Bellanca, J. (1995). *Designing professional development for change.* Arlington Heights, IL: IRI Skylight.

Berkeley Planning Associates (1997). *Classroom observation summary: Elementary and middle schools.* (Teacher Led Technology Challenge.) Oakland, CA: Author.

Brooks, J. G., & Brooks, M. G. (1999). *In search of understanding: The case for constructivist classrooms.* Alexandria, VA: Association for Supervision and Curriculum Development.

California Commission on Teacher Credentialing (2001). New Technology Standards, including a Plan for the Implementation of Programs to Meet the New Technology Standards—Coded Correspondence 99-9916. Retrieved June 1, 2001 from the World Wide Web: http://www.ctc.ca.gov/codcor.doc/999916/999916.html#20.5

CEO Forum on Education and Technology (2001). *School technology and readiness (StaR) chart.* Retrieved June 1, 2001 from the World Wide Web: http://www.ceoforum.org/reports.cfm?RID=3

CTAP (California Technology Assistance Program) (2001). *What is CTAP?* Retrieved June 1, 2001 from the World Wide Web: http://ctap.k12.ca.us/whatis.html

Danielson, C. (1996). *Enhancing professional practice: A framework for teaching.* Alexandria, Virginia: Association for Supervision and Curriculum Development.

Dodge, B. (2000). *Webquest rubric, version 1.02.* Retrieved June 1, 2001 from the World Wide Web: http://projects.edtech.sandi.net/staffdev/rubric.htm

Filippo, R. F., & Riccards, M. P. (2000). Initial teacher certification testing in Massachusetts: A case of the tail wagging the dog. *Kappan, 82*(1), 34-37.

Goldman, S., Cole, K., & Syer, C. (1999). *The technology/content dilemma.* Whitepaper presented at the 1999 Secretary's Conference on Educational Technology. Available online at: http://www.ed.gov/Technology/TechConf/1999/whitepapers/paper4.html

Grant, C. M. (1999). Beyond tool training: Professional development for elementary math teachers. *Learning and Leading with Technology 27*(3), 42-47.

Guiney, S. (1999). Technology funding in schools: Politics, pedagogy, and values. *School Business Affairs, 65*(8), 32-37.

Harvey, J., & Purnell, S. (Eds.). (1995). *Technology and teacher professional development* (OERI Publication No. DRLI-1045-CTI). Washington, DC: U.S. Department of Education.

Herr, P. (2000). The changing role of the teacher: How management systems help facilitate teaching. *T.H.E. Journal, 28*(4), 28-34.

Illinois State Board of Education (2001). *Division of Professional Preparation.* Retrieved June 1, 2001 from the World Wide Web: http://www.isbe.state.il.us/profdevelopment/Default.htm

ISTE (2001). National educational technology standards for teachers. Retrieved June 1, 2001 from the World Wide Web: http://cnets.iste.org/index3.html

Lemke, C. & Coughlin, E. (2001). *Technology in American schools: Seven dimensions for gauging progress.* Retrieved June 1, 2001 from the World Wide Web: http://www.mff.org/publications/publications.taf?page=158).

Louisiana Department of Education (1997). *Louisiana challenge grant.* Retrieved June 1, 2001 from the World Wide Web: (http://www.challenge.state.la.us)

Merriam-Webster (2001). *Merriam-Websert's Collegiate Dictionary.* Available online: http://www.m-w.com/netdict.htm

National Council of Teachers of Mathematics (1991). *Professional standards for teaching mathematics.* Reston, VA: Author.

National Research Council (1996). *National science education standards.* Washington, DC: National Academy Press.

Norman, M. M. (1999). Beyond hardware. *The American School Board Journal, 186*(7), 16-20.

Perry, S. (1998). Just what is a competency? (and why should you care?). *Training 35*(6), n.p. Available online: http://edweb.sdsu.edu/courses/ed791b/bmjjust.pdf

Roblyer, M. D., Edwards, J., & Havriluk, M. A. (1997). *Integrating educational technology into teaching.* Upper Saddle River, NJ: Merrill.

Rogers, E. M. (1995). *Diffusion of innovations* (4th ed.). New York: Free Press.

San Diego State University, Educational Technology Department (2000). *The Webquest page at San Diego State University.* Retrieved June 1, 2001 from the World Wide Web: http://edweb.sdsu.edu/webquest/

Sherwood, S. (1999). From computer lab to technology class. *Learning and Leading with Technology 27*(3), 28-31.

Simonson, M. R., & Thompson, A. (1997). *Educational computing foundations* (3rd ed.). Upper Saddle River, NJ: Merrill.

Sparks, D., & Hirsh, S. (1997). *A new vision for staff development.* Alexandria, Virginia: Association for Supervision and Curriculum Development.

Texas Classroom Teacher's Association (2001). SBEC approves certification standards and areas of teacher certification. Retrieved June 1, 2001 from the World Wide Web: http://www.tcta.org/cert/certstrc.htm).

Tomei, L. A. (1997). Instructional technology: Pedagogy for the future. *T.H.E. Journal*, *25*(5), n.p. Available online at: http://www.thejournal.com/magazine/vault/A1489.cfm

University of Kentucky, College of Education (2001). *50 state's certification requirements.* Retrieved June 1, 2001 from the World Wide Web: http://www.uky.edu/Education/TEP/usacert.html).

UTAP (Utah Technology Awareness Project) (2001). UEN news: Technology awareness project tutorial. Retrieved June 1, 2001 from the World Wide Web: http://wwwj1.uen.org/servlet/cgi/utap/utaphelp.cgi

5. Teacher Outcomes: Improved Technology Skills

Talbot Bielefeldt
International Society for Technology in Education

Introduction

The need for teachers to improve their comfort and skill level with computer technology has been recognized for many years in federal educational reports (Office of Technology Assessment, 1995). The National Center for Education Statistics notes that teachers' actual technology use is related to the adequacy of their technology preparation, but that only one-third of teachers feel well prepared to use technology (National Center for Education Statistics, 2000). Programs such as the Technology Innovation Challenge Grants (TICG), Star Schools, and Preparing Tomorrow's Teacher to Use Technology (PT^3) emphasize teacher technology training as a necessary condition of incorporating computers and other new tools into teaching and learning (Office of Educational Technology, 2000). Determining the level of teacher technology skills is thus an essential part of evaluating the effectiveness of the programs and the capacities and needs of educators whom the programs are designed to assist.

The assessment of teacher technology skills seems at first glance to be straightforward, at least in comparison to assessing other abilities teachers must have in classroom management, curriculum, content, and instructional design. Assuming a technology training program that follows recognized effective practice in adult learning, it would seem to be a relatively simple task to determine if teachers have acquired specific abilities to use particular applications and pieces of equipment. However, when we try to assess a particular teacher in a particular situation, the issues suddenly become more complicated. The complications stem from two main issues:

- It is not always clear which skills are important to assess, and to what extent. Instructional activities vary in the amount of technology skill they require of a teacher. Furthermore, teachers need to concentrate on different types of skills depending on what they have already learned about technology. Some are just mastering the basics of the hardware and software they want to use. Others are at a stage where the critical skill is knowing how to use technology in particular classroom settings or teaching situations (Office of Technology Assessment, 1995, pp. 134-135).

- Even when the most important skills to measure have been identified, it is not always clear *how* to assess those skills. Reliability, validity, feasibility, and other measurement issues apply to technology skills as much as to any other area of evaluation.

This chapter discusses these two issues in turn and examines how they apply to technology-skill evaluation instruments used by grantees in recent U.S. Department of Education programs.

What Do We Assess?

The Issue Of Relevance

One might suggest that everyone in any computer-using society should probably know at least word processing. However, word processing now encompasses many kinds of document manipulation. The range extends from simple keyboarding to producing multimedia presentations that incorporate graphics, video, sound, and interactive hyperlinked text. The question of where we draw the line at what is "just word processing" can only be answered in terms of who is using the technology, and for what. Seymour Papert, the developer of the Logo programming language, wrote that the only important computer skill is "the skill and habit of using the computer in doing whatever one is doing" (Papert, 1995). In our example, word processing requirements might be different for an English teacher, the faculty sponsor of the student newspaper, an art teacher, and a mathematician.

This issue has cropped up in previous decades in relation to students' "computer literacy"—the basic, general-purpose skills that would enable them to function around digital technology in the same way that reading and writing enable students to work with written media. For many educators, such as Papert, cited above, the term "computer literacy" has assumed a derogatory connotation based on the experience that computer-literacy classes often have not provided students with the skills they need to actually use computers in their school work. There is data to suggest that the same situation exists in pre-service teacher education. In a survey of teacher education institutions in the U.S., Moursund and Bielefeldt (1999) found that for most programs, there is little correlation between number of technology-specific course credits required and other measures of capacity, such as technology skills of pre-service teachers or integration of technology into other college coursework.

Defining Relevant Technology Skills

Standards

In recent years there have been several attempts to establish general standards that would help institutions decide what technology skills are important for teachers to acquire. The CEO Forum offers its School Technology and Readiness (STaR) Charts for schools and teacher training institutions. The STaR Chart uses stages of skill development (Entry/Adoption/Adaptation/Appropriation/Invention) identified in the Apple Classrooms of Tomorrow (ACOT) in the 1980s and early 90s. The chart for "Pattern of Teacher Technology Use" (CEO Forum on Education and Technology, 1997) appears as Table 1:

Table 1. CEO Forum School Technology and Readiness Chart: Integration and Use

Stage and ACOT level	Pattern of Teacher Technology Use
Low-Tech: Entry and Adoption Skill Stages	None. (Computers in labs run by computer instructors.)
Mid-Tech: Adaptation Level Skill Stage	Regular use by some teachers. Some word processing to construct assignments and tests. Limited use of Internet to access ideas for curriculum.
High-Tech: Appropriation Level Skill Stage	Strong grasp of multimedia software and regular use of online resources. Some teachers take professional development courses or join peer discussion groups online. Regular use of word processing.
Target-Tech: Invention Level Skill Stage	Regular use of technology to access remote information, communicate with students and parents, and complete administrative tasks such as student progress repots, databases and word processing.

The STaR chart is designed as an institutional needs assessment rather than an individual skill assessment, and the levels obviously describe multiple complex behaviors for a group. "Teacher technology use" is defined not just by knowing what keys to press, but by a pattern of professional development and integration with curriculum and professional tasks.

A more detailed and individually-oriented approach is the set of standards introduced in North Carolina in 1995. North Carolina's Technology Competencies for Educators (North Carolina Department of Public Instruction, 1995) are extremely specific, with 114 indicators at basic and advanced levels. The distinction between the levels is that the basic competencies emphasize technology skills alone, whereas the advanced levels emphasize the use of technology in teaching and learning. An excerpt from the North Carolina competencies appears in Table 2, with a sample of basic and more developed skills presented under each area of competency.

A more recent standard is the Milken Family Foundation's Professional Competency Continuum for Core Technology Skills (Coughlin & Lemke, 1999). Other professional competencies developed by Milken for educational technology include administration, classroom and instructional management; curriculum, learning, and assessment; and professional practice. Professional competency is in turn one of Milken's "Seven Dimensions for Gauging Progress" in educational technology, a model that also includes learners, the learning environment, system capacity, community connections, technology capacity, and accountability.

As with the CEO Forum STaR Chart, Milken aligns teacher skills with stages of adoption (Entry/Adaptation/Transformation). Milken's assessment tool asks teachers to rate themselves (or administrators to rate teachers) on a scale of 1-10 from "entry" to "transformation" for 25 competencies. The topics of questions include hardware/computer, hardware/other, applications, information tools, network tools, and multimedia/presentation tools (see Table 3).

Another set of detailed standards is that developed by the International Society for Technology in Education (2000). Originally created as part of the accreditation process for the National Council for Accreditation of Teacher Education (NCATE), the standards were revised under a 1999 U.S. Department of Education PT3 grant. Now dubbed the National Educational Technology Standards (NETS) for Teachers, they are aligned with a related set of NETS for students.

Technology skills *per se* are included in NETS under the first standard, Technology Operations and Concepts. (Other standard types include (II) Planning and Designing Learning Environments and Experiences; (III) Teaching, Learning, and the Curriculum; (IV) Assessment and Evaluation; (V) Productivity and Professional Practice; and (VI) Social, Ethical, Legal, and Human Issues. Standard I is addressed by 21 of 77 "performance profile" indicators at four levels: general preparation, preservice training, student teaching, and in the first year of teaching in a school. Table 4 shows the first four indicators for Technology Operations and Concepts at each career level. The Roman numerals following each indicator refer to the standard which it addresses.

Table 2. North Carolina Technology Competencies for Educators (Excerpts)

Basic	Advanced
1.0 COMPUTER OPERATION SKILLS 1.1 Start up and shut down computer system and peripherals 1.17 Install/reinstall and update system software and printer drivers 2.0 SETUP, MAINTENANCE, AND TROUBLE-SHOOTING 2.1 Setup computer system and connect peripheral devices 2.8 Protection against computer viruses 3.0 WORD PROCESSING/INTRODUCTORY DESKTOP PUBLISHING 3.1 Enter and edit text and copy and move a block of text 3.8 Insert clip art into document 4.0 SPREADSHEET/GRAPHING 4.2 Enter data in an existing spreadsheet 4.6 Insert a spreadsheet into a word processing document 5.0 DATABASE 5.2 Sort a database by specific fields, add and delete record 5.5 Insert database fields into word processing document 6.0 NETWORKING 6.1 Use a file server (connect/log on, retrieve a program or document, save a document to a specified location) 6.4 Select/de-select a network zone 7.0 TELECOMMUNICATIONS 7.1 Connect to the Internet or an on-line service 7.16 Use effectively distance learning, desktop video conferencing, and teleteaching technologies 8.0 MEDIA COMMUNICATIONS 8.1 Produce print-based products (e.g., newsletters, brochures, posters, books) 8.12 Produce a video 9.0 MULTIMEDIA INTEGRATION 9.1 Use a linear multimedia presentation 9.7 Input and digitize sound from microphone and audiocassette player/recorder	10.0 CURRICULUM 10.1 Use the Computer Skills Curriculum to identify what students should know/be able to do 10.6 Locate, evaluate, and select appropriate teaching/learning resources and curriculum materials for the content area and target audience, including computer-based products, videotapes and discs, local experts, primary documents and artifacts, texts, reference books, literature, and other print sources 11.0 SUBJECT-SPECIFIC KNOWLEDGE 11.1 Use technology in the discipline/subject for learning and as a medium for communications 11.4 Use technology to facilitate teaching strategies specific to the discipline 12.0 DESIGN AND MANAGEMENT OF LEARNING ENVIRONMENTS/RESOURCES 12.1 Develop performance tasks that require students to (a) locate and analyze information as well as draw conclusions and (b) use a variety of media to communicate results clearly 12.7 Select and create learning experiences that are appropriate for curriculum goals, relevant to learners, based upon principles of effective teaching and learning, incorporate the use of media and technology for teaching where appropriate, and support learner expression in a variety of media using a variety of media communication tools 13.0 CHILD DEVELOPMENT, LEARNING AND DIVERSITY 13.1 Use media and technology to address differences in children's learning and performance 13.4 Use appropriate local, state, and national services or resources to meet diverse learning needs through technology 14.0 SOCIAL, LEGAL, AND ETHICAL ISSUES 14.1 Establish classroom policies and procedures that ensure compliance with copyright law, fair-use guidelines, security, and child protection 14.3 Social, legal, and ethical issues related to technology use

Table 3. Milken Family Foundation Professional Competency Continuum: Core Technology Skills (excerpt)

Entry	**Adaptation**	**Transformation**
Educators, students, and the community are aware of the possibilities, yet learning, teaching and the system remain relatively unchanged by technology. <--Lower	*Technology is thoroughly integrated into existing practice.*	*Technology is a catalyst for significant changes in learning practices.* Higher-->
O O O	O O O O	O O O
Operate the computer at a novice level (I can use folders and basic menu commands but cannot change settings or use file management capabilities).	*Operate the computer at an intermediate level (I can use the file management capabilities of my operating system; change basic settings such as colors and mouse speed; and connect simple peripherals such as mouse and keyboard).*	*Operate the computer at an expert level (I can operate all aspects of the operating system without effort).*
Start, quit, and use basic functions of a word processor, but do not use other applications in a basic suite such as database or spreadsheet.	*Use the basic functions of a suite of office software programs such as Appleworks or Microsoft Office.*	*Use basic and advanced functions of a suite of office software programs such as Appleworks or Microsoft Office and can move data between database, word processor and spreadsheets.*
Obtain assistance if my computer breaks down but I cannot fix any problems	*Recognize common problems with my computer, can explain what is wrong to the repair person, and fix minor problems.*	*Fix most computer problems that are user serviceable (e.g., change network settings.)*
Recognize when my computer is inadequate to the task but am not comfortable recommending any kind of upgrade or enhancement.	*Recommend upgrades to make a computer system work better (e.g., memory expansion).*	*Understand when upgrades are needed, can make recommendations regarding upgrades and may serve as a resource to others.*
Learn new features of a computer system but I need the assistance of another person.	*Learn new features of a computer system independently or with minimal assistance.*	*Teach myself just new features of a computer system. I sometimes serve as a resource to others.*

Note that some indicators address multiple standards. Only three indicators address the Technology Operations and Concepts standard alone. In other words, the NETS indicators imply that assessment of teacher technology competencies should not be viewed in isolation, but should focus on how well teachers integrate technology into learning.

Table 4. NETS Technology Performance Profiles for Teacher Preparation (excerpt)

Upon completion of the general preparation component of their program, prospective teachers:	**Prior to the culminating student teaching or internship experience, prospective teachers:**	**Upon completion of the culminating student teaching or internship experience, and at the point of initial licensure, teachers:**	**Upon completion of the first year of teaching, teachers:**
*1. demonstrate a sound understanding of the nature an operation of technology systems. (I)** *2. demonstrate proficiency in the use of common input and output devices; solve routine hardware and software problems; and make informed choices about technology systems, resources, and services. (I)** *3. use technology tools and information resources to increase productivity, promote creativity, and facilitate academic learning. (I, III, IV, V)*	*1. identify the benefits of technology to maximize student learning and facilitate higher order thinking skills. (I, III)* *3. identify technology resources available in schools and analyze how accessibility to those resources affects planning for instruction. (I, II)* *4. identify, select, and use hardware and software technology resources specially designed for use by PK-12 students to meet specific teaching and learning objectives. (I, II)*	*1. apply troubleshooting strategies for solving routine hardware and software problems that occur in the classroom. (I)*	*1. assess the availability of technology resources at the school site, plan activities that integrate available resources, and develop a method for obtaining the additional necessary software and hardware to support the specific learning needs of students in the classroom. (I, II, IV)* *2. make appropriate choices about technology systems, resources, and services that are aligned with district and state standards. (I, II)* *6. plan for, implement, and evaluate the management of student use of technology resources as part of classroom operations and in specialized instructional situations. (I, II, III, IV)*

The ISTE standards are cited in the definition of several state standards, including those of Kentucky (Office of Education Technology, 2001), Idaho (Educational Technology Assessment Office, 2000), and Louisiana (Louisiana Center for Educational Technology, 1999). The Idaho teacher technology proficiency testing program is an example of intermingled integration and technology skills. Idaho offers a multidimensional approach in which teachers have a choice of being evaluated by a multiple choice test (e.g., “Using a drill and practice program is most appropriate for (a) students learning new concepts; (b) students practicing concepts they have learned; (c) simulations of real world activities”) or by performance assessment (e.g., “find a World Wide Web site and review it in an e-mail document”). While the multiple-choice test also includes keystroke-specific skill questions, the advertised philosophy behind the performance assessment is that students need to develop skills in context:

> The software listed in the competencies are the standard tools used by business, industry, and science to record, store, transmit, and manipulate information to solve a problem or to produce a product. . . . Although these tools are sometimes taught as classes by themselves, it is the unanimous opinion of the educators that produced this test that their most appropriate use is to speed or deepen student learning in the content areas. There is no value in teaching a student to use a word processor, the Internet, etc., unless the student learns to apply the tool to solve a problem or produce a product (Educational Technology Assessment Office, 2000).

Item Selection

A second type of guide to what is important to assess in teacher technology skills comes from the evaluation tools used in technology initiatives. Evaluation instruments used in Technology Innovation Challenge Grant (TICG) programs tend to mix direct questions about technology skills *per se* with context questions about classroom integration. For instance, teacher surveys used in the ACT Now! TICG in Chula Vista, CA (Table 5) asked teachers to rate themselves on scales regarding their skill or comfort with various applications. Separate rating scales address frequency of use and whether the applications were used in or outside of the classroom (Bober, Harrison, & Lynch, 1997).

Table 5. Teacher Survey, ACT Now! (excerpts)

<table>
<tr><td colspan="2">1. Please use the following scale to describe your skills in using the types of applications/software listed.
1 = no experience
2 = beginner (have used but consider myself a beginner)
3 = intermediate (use it routinely but not to its full capacity)
4 = advanced (use it routinely to its full capacity)
5 = mastery (could offer training to others)</td><td colspan="2">2. Please use the scale that follows to describe the extent to which you use various applications/software to support your classroom instruction. (Check both 2 and 3 if appropriate.)
1 = do not use
2 = use it myself to prepare for classes
3 = use instructionally with students</td></tr>
<tr><td>a. Word processing (e.g., Word)</td><td></td><td>a. Word processing (e.g., Word)</td><td></td></tr>
<tr><td>b. Presentation (e.g., PowerPoint)</td><td></td><td>b. Presentation (e.g., PowerPoint)</td><td></td></tr>
<tr><td>c. Multimedia authoring (e.g., HyperStudio)</td><td></td><td>c. Multimedia authoring (e.g., HyperStudio)</td><td></td></tr>
<tr><td>d. Classroom management (e.g., lesson planning, grading)</td><td></td><td>d. Classroom management (e.g., lesson planning, grading)</td><td></td></tr>
</table>

An annual Teacher Technology Survey for **Global Connections** in Phoenix, Arizona, pursues a similar strategy (Bielefeldt, 2000). For each of 19 applications used by the Phoenix Union High School District, teachers complete four scale items addressing frequency of use in and out of class, frequency of access, and perceived level of skill. (Table 6).

Table 6. Excerpt, Global Connections Teacher Technology Survey

On average during this school year, how often did you...	...use the following hardware and software outside the classroom?	...use the following hardware and software in the classroom w/students?	...have access to the following hardware and software?	How comfortable are you in using the following hardware and software? 1=do not know this technology 2=need assistance 3=can perform basic operations 4=can perform advanced operations 5=can teach this technology
	1=never 2=a few times 3=once a week 4=once a month 5=every day			
Presentation System				
CD-ROM/DVD				
Digital Camera/Flexcam				
Scanner				
Video camera/ Camcorder				

Some survey instruments mix technology use and integration items on the same scales; see for instance the items excerpted in Table 7 from the Teacher LED Technology Challenge TICG questionnaire by Toms Barker (1999b).

Table 7. Teacher LED Technology Challenge
Follow-Up Teacher Questionnaire (Excerpt)

For each of the following, please indicate the level of skill that is closest to your level				
No experience	Beginner	Literate, comfortable	Independent, fluent	Expert, can innovate
Using computers for word processing.				
Using computers to produce graphics or art.				
Using computers and other technology to create multimedia products.				
Using computers for spreadsheets, record keeping, and data processing.				
Selecting educational software to match a particular curriculum goal.				
Integrating the use of the computer into regular instructional activities.				

Of a dozen TICG teacher evaluation surveys and observation protocols reviewed for this chapter, all addressed both teacher technology skill and ability to integrate technology into professional activities. Standards and assessment instruments might segregate keystroke-level skills from integration skills in different dimensions or scales, but at the time of data collection, both aspects of technology use are addressed at once.

Examples from Star Schools and Challenge Grant Evaluations

Surveys

The other major issue is the actual technique for "getting at" the skills we want to assess. The most common type of instrument used in U.S. Department of Education technology programs is the self-report survey. Surveys have obvious advantages: They can be used to collect data from large numbers of people in a short period of time. If they employ scales or other quantitative response modes, the data can be quickly assembled and analyzed with the aid of database and statistical software. The same software can be used to determine the reliability of individual items and of the entire instrument.

This information is useful for adopting and adapting instruments to different settings. For instance, the Technology Proficiency Self-Assessment (TPSA), a survey used with the **Key Instructional Design Strategies (KIDS)** Challenge Grant in Allen, Texas, is made up of 20 statements about technology use, grouped into four domains (five questions each)—e-mail, the Web, integrated applications, and teaching with technology (Table 8). From the standpoint of someone who may want to adopt the survey in an evaluation, it is useful to know that the reliability of the instrument is quite high (α=.94), as are the reliabilities of the subscales in each domain (α=.78 or greater for each) (Knezek, Christensen, Miyashita, & Ropp, 2000). This suggests that the separate question groups might be used as mini-assessments when the whole survey is not appropriate.

Table 8. Technology Proficiency Self-Assessment

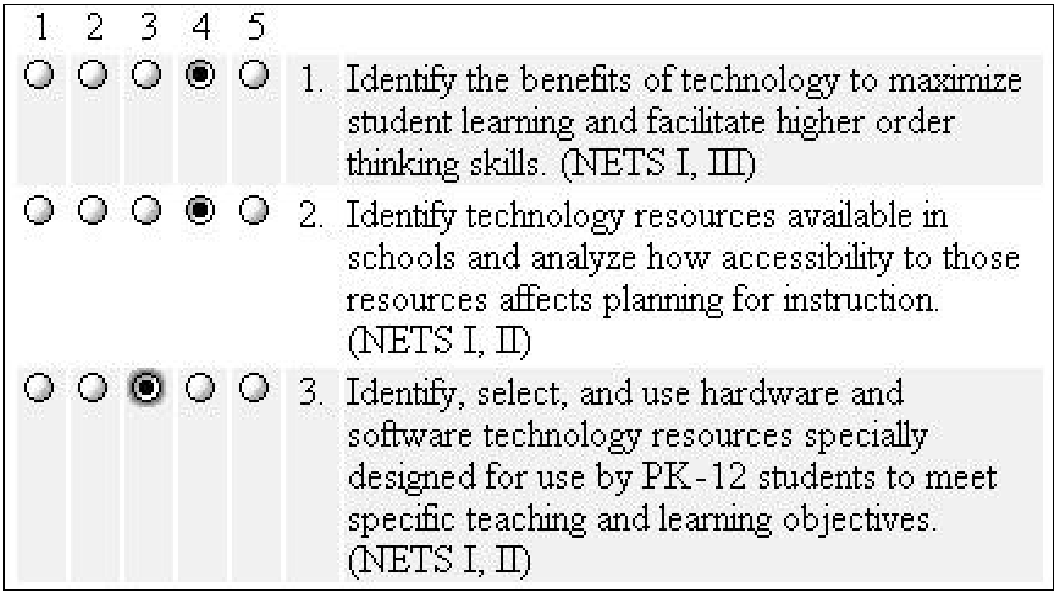
Select one level of agreement for each statement to indicate how you feel.
SD = Strongly Disagree, D = Disagree, U = Undecided, A = Agree, SA = Strongly Agree

I feel confident that I could …

1. send email to a friend.
2. subscribe to a discussion list.
3. create "nickname" or an "alias" to send e-mail to several people.
4. send a document as an attachment to an e-mail message.
5. keep copies of outgoing messages that I send to others.
6. use an Internet search engine (e.g. Infoseek or Alta Vista) to find Web pages related to my subject matter interests.
6. search for and find the Smithsonian Institution Web site.
7. create my own World Wide Web home page.
8. keep track of Web sites I have visited so that I can return to them later.
9. find primary sources of information on the Internet that I can use in my teaching.
10. use a spreadsheet to create a pie chart of the proportions of the different colors of M & M's in a bag.
11. create a newsletter with graphics and text in 3 columns.
12. save documents in formats so that others can read them if they have different word processing programs.
13. use the computer to create a slideshow presentation.
14. create a database of information about important authors in a subject matter field.
15. write an essay describing how I would use technology in my classroom.
16. create a lesson or unit that incorporates subject matter software as an integral part.
17. use technology to collaborate with other interns, teachers, or students who are distant from my classroom.
18. describe 5 software programs that I would use in my teaching.
19. write a plan with a budget to buy technology for my classroom.

Like many newer instruments, this one is administered online. Technology has enhanced the appeal of surveys because of the ability to offer the instruments as interactive Web forms, with data entry and even initial analysis available "on the fly" as soon as a respondent completes the form. A survey tool used in several federal technology programs is Profiler, a Web application developed at the South Central (now High Plains) Regional Technology in Education Consortium (HPR*TEC, 2000). With Profiler running on a Web server, administrators or technology coordinators can design a survey and offer it to a group—teachers in a building or district, for instance, with password protection to control access. A Profiler Technology Literate Teachers survey created for the NETS standards asks respondents to estimate their competence level for each of 24 NETS indicators, using a five-point scale: (1) unable, (2) basic, (3) developing, (4) proficient, and (5) advanced. (See Figure 1).

Figure 1. Profiler screen capture.



Each individual's submission is displayed on a radial graph (Figure 2), along with a companion graph for the group averages. The sections around the border indicate groups of questions defined by the survey authors. The graph as displayed by the Profiler server is "hot": moving the mouse over a question displays the item and the respondent's score.

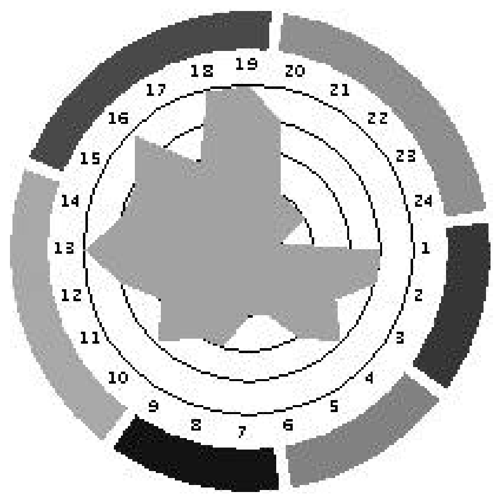


Figure 2 Profiler Radial Graph

Despite their advantages and popularity, self-report surveys have obvious drawbacks: Individuals are not always the best judges of their own skill levels. Responses may be affected by incomplete knowledge of the field, inaccurate self-perceptions, or social-desirability pressures (to appear modest, or competent, etc.). Although self-report instruments may provide accurate measures of *perceptions about* technology skills, they do not measure the skills themselves. Research by Ropp (1999) on the TSPA and other instruments suggests that the relation between attitudes toward technology and skills with technology is tenuous.

Tests, Observations, and Products

There are various approaches in use that are more or less objective. "Objective" in this sense means that the assessment data are products or events that can be observed. This means that—in theory, given enough time and resources—they can be simultaneously or repeatedly observed by any number of observers, and the results compared to establish the reliability and validity of the data.

The data may be formal examinations, as in the Idaho proficiency tests, or they may be analyses of work samples, as in a rubric used in the **Education Future NOW** project in North Carolina to assess technology-infused lesson plans (Schulze & Holland, 2000). The Education Future NOW rubric has four parts, worth a total of 100 points: The first part focuses on teacher multimedia skills; the other three address the organization of lessons, student projects, and student assessment. The multimedia section, reproduced as Table 9, consists of a list of specific criteria relating to the application used and product produced:

Table 9. Education Future NOW Technology
Integrated Unit of Study Rubric (excerpt)

Part I - Multimedia presentation (25 points)
Create a PowerPoint, HyperStudio or KidPix presentation to introduce or illustrate the key concepts in the unit.
__Standard Course of Study Goal
__Minimum of 10 slides
__At least five of the following items: clip art, digital pictures, scanned pictures, Internet information, tables, charts, sound, voice recordings

Another approach to objectivity is to survey a second party about past observations of technology skills. Examples of this approach are the **Eiffel Project** surveys of technology coordinators and administrators (Denes, 1998), and the ISTE survey of Information Technology in Teacher Education (Moursund & Bielefeldt, 1999), which was completed mainly by education college deans and department chairs. The first instrument asks respondents to rate teacher's skills with various applications on a five-point scale ("none," "low," "adequate," "good," and "excellent"). The second survey asks administrators to estimate the percentage of student teachers in their programs who can demonstrate basic skills with word processing, e-mail, and web browsing.

The type of assessment most anchored in context is direct observation of the learning environment. The virtue of observations is their face validity. There is no need to infer how the technology skill might play out in the classroom. Of course, observers may disagree on the interpretation of what they see—is a particular PowerPoint presentation true integration of technology into the curriculum, or just the teacher demonstrating a knowledge of the software? However, the observations can be repeated to establish average levels of behavior. With the aid of technology (video), a single observation can be recorded and archived. Competing interpretations can be argued out over time as the observation data is reanalyzed.

Two general approaches to collecting classroom observation data are represented in the instruments reviewed for this chapter. The first approach is narrative, in which observers are asked to describe technology use as they see it, and in their own words. The other approach is a checklist, in which observers record or rate behavior in predefined categories. Instruments usually have both types of response options, but emphasize one or the other. An example of a more narrative approach is the Classroom Observation Summary from the **Teacher-Led Technology Challenge** (Toms Barker, 1997a). The "Technical Issues" section asks for an overall rating of how proficient the teacher is in using technology ("very," "somewhat," "slightly," or "not at all"), with several lines in which to explain the rating. Then it offers a large area in which to make notes on technical help offered, problems that arose, and solutions to the problems.

A checklist approach is exemplified by the Classroom Observation Instrument developed by WestEd and used in The **Connections Project** in Nebraska (Timms, 1999). This instrument calls for observers to fill out eight ratings and categorical classifications every five minutes for 45 minutes. The focus of this instrument is primarily the integration of technology, and most of the items have to do with elements of pedagogy. The one technology item asks the observer to rate uses of technology and to describe *how* computers were used (Table 10).

Table 10. Excerpt, The Connections Project Classroom Observation Instrument

Use of Technology **Minutes:**	**5**	**10**	**15**	**etc.**	**How computers were used (check all that apply)**
1. Not used.	1	1	1	...	1. Text 2. Numeric 3. Graphic
2. Add-on	2	2	2	...	4. Research 5. Communication
3. One-dimension infusion	3	3	3	...	6. Other (specify)
4. Multi-dimension infusion	4	4	4	...	Describe hardware/software:

The terms are defined in an observer's manual that is provided with the response form. (For example, "2. Add-on: Limited use of computer technology by students and teacher. The use of technology is simplistic, not well integrated into the lesson, and does not support learning in a meaningful way.")

The advantage of such a checklist is that it allows an observer to quantify certain characteristics of a very complex environment. With the addition of the fixed-interval observations, the data includes a record of the extent of the behaviors observed. It minimizes the risk that an observer will miss instances of a behavior, or will give inordinate weight to a single instance that is for some reason memorable (such as having occurred near the end of the observation).This approach works by directing the observer to pay attention to only certain features of the environment and only at certain times.

A general point is that all of these measures are only relatively "objective." Compared with self-report surveys, the subjectivity of the teachers whose technology use is of interest is removed. On the other hand, the subjectivity of the instrument designers and of the individuals observing the teachers is still very much an issue. The reliability and validity of assessments always need to be confirmed by replication, alternate measures, and outcomes.

Discussion

The issues in assessing teacher technology skills might best be seen as a set of tradeoffs. Every evaluation instrument and set of standards has to implicitly or explicitly address these issues, and their resolution for each program is central to the evaluator's role.

General Standards vs. Specific Conditions

Standards are a common feature across the educational landscape. The challenge for educators and educational evaluators is translating general standards to the wide variety of situations encountered in the field. The existence of standards as a condition of evaluation essentially dictates a goal-based evaluation; that is, goals (representing standards) and assessments must be aligned. A misalignment means that results may not be accepted or not understood by the evaluation audience.

Choosing data and assessment techniques that relate to the standards is the obvious response to the challenge, but in situations where the evaluator has a role in shaping a program, it may be also appropriate to choose the standards that best fit the context. As we noted above, some standards (e.g., the Milken Professional Competencies) delineate between technology skills and integration skills, while others (e.g., NETS) present technology skills in the context of integration. The Milken Competencies may offer more detailed examples upon which to base assessments of new teachers just learning hardware and software skills. The NETS standards may be a more meaningful target for teachers who are moving beyond basic skills into integration.

It is also important to note the scope and implicit assumptions of each set of standards. For instance, more than half of the NETS teacher indicators are derived from the NETS *student* standards (International Society for Technology in Education, 1998), and are supposed to be met during the "General Preparation" period prior to beginning a teacher education program. That is, the NETS writers clearly anticipate an environment in which many of these competencies will be acquired and assessed while teachers-to-be are still students in high school or in lower-division college programs.

Technology Skills vs. Technology Integration

As noted earlier, evaluation practice clearly supports assessing skills along with or in the context of integration. However, another assumption in many standards (e.g., the North Carolina Technology Competencies) and assessments (e.g., The Connections Project Classroom Observation Instrument) is that teachers will move from basic skills to integration as their experience increases. We have seen skill-vs.-integration status used to define the stage of development, but in cases where the evaluator needs to make a summative evaluation, it is important to know the teacher's experience *before* making a judgment. Otherwise, individuals and programs may be held to the wrong standards. Questions about context (extent of training, years of teaching, access to technology) should be part of any skill assessment. Furthermore, experienced technology users may become novices again overnight as they move into learning new software or hardware, so labels such as "technology-proficient" need to be constantly checked for accuracy.

Self-Report vs. Objective Measures

The distinction between surveys and objective measures was discussed earlier as a tradeoff between validity and feasibility. However, the choice of instrument also has to do with the stage of the evaluation. Self-reported use of technology recorded in surveys or interviews may be the most relevant variable in a professional development needs assessment. Objective observation of technology use might be the preferred strategy for evaluating the outcomes of that professional development.

The quality of implementation of each approach also determines its ultimate value. A survey that is well grounded in context and for which reliability data is available may be more useful than an "objective" measure that is inconsistently applied. Unfortunately, many of the instruments reviewed for this chapter do not have well-documented guidelines for use. One exception in the survey family is the Technology

Proficiency Self-Assessment, which appears in a manual from the University of North Texas that includes administration instructions, reliabilities, and analysis approaches (Knezek, Christensen, Miyashita, & Ropp, 2000). An observation tool that is published with coding guidelines is The Connections Project Classroom Observation Instrument. Evaluators also report that observers underwent pre-observation training to ensure inter-rater reliability (R. Pawloski, personal communication, October 14, 2000).

A special note is due the second-party survey, in which an administrator or other authority is asked to assess teachers with whom they work. The author has pointed out in reference to one of his own instruments (Moursund & Bielefeldt, 1999, p. 11) that this technique, while removing the subjectivity of the teacher who is using the technology, relies heavily on the subjective opinion of the respondent. The main benefit of this approach is that a single respondent can provide estimates of the behavior of a number of subjects, and thus it allows coverage of a large population in a short period of time. However, it is only appropriate where the quality of the respondents' knowledge is already well established.

Definition of Terms

Finally, an issue common to all assessments of technology skills is defining exactly what aspects of technology use are being described. As we have seen, text-based communication with computers may be represented as generally "word processing," or as specifically as "create a newsletter with graphics and text in three columns," or "enter and edit text and copy and move a block of text."

Greater specificity is an obvious advantage when evaluators want to aggregate results from different programs. It is difficult to say if "word processing" improves student writing if we do not have a common definition of the independent variable. Evaluators can clarify their meaning either by writing detailed item descriptions or by writing more items that break down behavior into easily identifiable components. Either strategy runs the risk of producing instruments that are lengthy or cumbersome to use.

Standards play a role here, in that they define the concepts or behaviors or interest in advance. However, standards only help to the extent that people agree to subscribe to them and interpret them in similar ways. Ultimately then, the issue of definition is one of communication between evaluators, programs, and funding agencies about the aspects of technology skill that are essential. Hopefully, resources such as this sourcebook will facilitate that effort.

References

Bielefeldt, T. (2000). *Global Connections technology survey. Global Connections.* Portland, OR: International Society for Technology in Education.

Bober, M., Harrison, P. J., & Lynch, E. W. (1997). *Teacher technology survey. ACT Now!* . San Diego, CA: San Diego State University.

CEO Forum on Education and Technology. (1997). *School technology and readiness report: From pillars to progress*. Washington, DC: Author.

Coughlin, E. & Lemke, C. (1999). *A continuum for professional development: Gauging professional competency for technology-supported learning.* Santa Monica, CA: Milken Family Foundation. Available online: http://www.mff.org/publications/

Denes, V. (1998). *Technology coordinator questionnaire and principal/director questionnaire*. (Eiffel Project.) New York: Columbia University, Teachers College, NCREST—National Center for Restructuring Education, Schools, and Teaching.

Educational Technology Assessment Office. (2000). *Idaho technology competency examination.* Boise, ID: Boise State University. [Online]. Available: http://education.boisestate.edu/itce/

High Plains Regional Technology in Education Consortium. (2000). *Profiler online collaboration tool.* Available: http://profiler.hprtec.org

International Society for Technology in Education. (1998). *National educational technology standards for students*. Eugene, OR: Author. [Online]. Available: http://cnets.iste.org/sfors.html

International Society for Technology in Education. (2000). *National educational technology standards for teachers*. Eugene, OR: Author.[Online]. Available: http://cnets.iste.org/perfprofiles.html

Knezek, G., Christensen, R., Miyashita, K., & Ropp, M. (2000). *Instruments for assessing educator progress in technology integration*. Denton, TX: University of North Texas.

Louisiana Center for Educational Technology, Louisiana Department of Education. (1999). *K-12 educational technology standards.* Available online: http://www.doe.state.la.us/DOE/asps/home.asp?I=LCET.

Moursund, D. & Bielefeldt, T. (1999). *Will new teachers be prepared to teach in a digital age?* Santa Monica, CA: Milken Family Foundation.

National Center for Educational Statistics. (2000). *Teachers' tools for the 21st century*. Washington, DC: U. S. Department of Education, Office of Educational Research & Improvement. Available online: http://nces.ed.gov/pubs2000/2000102.pdf.

North Carolina Department of Public Instruction. (1995). *North Carolina technology competencies for educators.* Available online: http://www.ncpublicschools.org/tap/techcomp.htm

Office of Education Technology, Kentucky Department of Education. (2001). KETS master plan for education technology FY2001-FY2006. Frankfort, KY: Author. [Online.] Available: http://www.kde.state.ky.us/oet/planning/mpupdate2001/

Office of Educational Technology, U.S. Department of Education. (2000). *Technology grant programs.* Available online: http://www.ed.gov/Technology/edgrants.html

Office of Technology Assessment, U.S. Congress. (1995). *Teachers and technology: Making the connection.* Washington, DC: U.S. Government Printing Office.

Papert, S. (1995). *The children's machine.* New York: Basic Books.

Ropp, M. (1999). Exploring individual characteristics associated with learning to use computers in preservice teacher preparation. *Journal of Research on Computing in Education, 31*(4), 402-424.

Schulze, S. & Holland, B. (2000). *Technology integrated unit of study rubric.* Education Future Now . North Carolina School of Science and Mathematics.

Timms, M. (1999). Classroom observation instrument. Connections Project . WestEd.

Toms Barker, L. (1997a). *Classroom observation summary: Teacher Led Technology Challenge.* Oakland, CA: Berkeley Planning Associates.

Toms Barker, L. (1997b). *Follow-up teacher questionnaire: Teacher Led Technology Challenge.* . Oakland, CA: Berkeley Planning Associates.

6. Technology Integration

Mario Yepes-Baraya
Harcourt Educational Measurement

Introduction

This chapter discusses the measurement of technology integration in schools in the context of the Technology Innovation Challenge Grants (TICG) and the Star Schools Program (SSP), sponsored by the Office of Educational Research and Improvement (OERI) of the U.S. Department of Education.

The chapter first describes OERI's vision for educational technology. Second, some basic definitions and assumptions are discussed. Third, the chapter presents two approaches to conceptualizing technology integration, including a framework based on learning theories and educational practice and a logic model for planning and evaluating educational interventions. Finally, a review of measures of technology integration extracted from instruments currently being used in various Challenge and Star Schools projects is provided. The framework and model presented earlier are used to organize and interpret the measures reviewed.

OERI's Vision for Educational Technology

OERI's vision for educational technology can be summarized as follows (U.S. Department of Education [U.S. DOE], 1999)[1]:

- To integrate current and future information and communication technologies into *everyday* teaching and learning;
- To create programs that capture the *opportunities* promised by emerging technologies in ways that are *affordable*, *sustainable*, and *scalable*; and
- To reform the *context* in which technology is being integrated.

This vision is predicated on two key elements: school reform and the opportunities afforded by integrating emerging technology. School reform refers to coherent approaches to school improvement that address curriculum and instruction,

[1] Emphasis added.

student assessment, teacher professional development, parent involvement, and school management. Unlike previous technologies, emerging computer-based technology, because of its interactivity, speed, access to multiple sources of information, and potential for cost effectiveness, offers numerous opportunities to achieve the goals proposed by school reformers. The joint implementation of school reform practices and emerging technology should enable educators to:

- Customize and improve teaching;
- Individualize and improve learning;
- Achieve equity; and
- Effect systemic changes (U.S. DOE, 1999).

Why Technology Integration is Important

Technology is rapidly becoming a dominant force in U.S. schools. Technology is regarded by many as a key element of education reform (U.S. DOE, 1999; McNabb, Hawkes, & Rouk, 1999). Although there is a growing body of research suggesting that technology can enhance and improve learning (Jonassen & Reeves, 1996), there is no consensus as to how this can be accomplished. It is generally recognized, however, that technology by itself is not sufficient to achieve the types of changes envisioned by reformers. In addition to technological tools, the DOE has recognized that educational reforms are needed that support improved teaching and learning:

> Learning technologies are effective only when treated as one component in implementation strategies that also encompass (1) curricular reform, (2) sophisticated and multiple assessments, (3) effective professional development, (4) well-maintained technology infrastructures and support systems, (5) attention to equity, and (6) the restructuring of organizational practices, budgets, and policies.
>
> An exemplary technology-based program will incorporate all of these dimensions. A promising one may incorporate some of them and will include plans to achieve the remainder. The ultimate goal of the linked elements of any technology program …is increased student learning (U.S. DOE, 1999).

Although some may not agree that *all* of the above elements are necessary to improve learning, the effectiveness of technology will clearly vary with situational variables, such as the goals and resources for instruction, the cognitive demands on the learner, the extent to which all learners' needs are considered, the teachers' comfort and skill with technology, and, of course, the types of technology available. Thus, "*integrated technology" is technology that supports and enhances the achievement of specific teaching and learning goals*. For example, if these goals include the mastery of basic arithmetic skills or the development of effective critical thinking skills, the criteria to assess the integration of the technology tools used should include the extent to which these tools contribute to the attainment of these goals in more effective and efficient ways than if more traditional means were utilized.

Learning Theories, Educational Context and Technology Integration

The Cognition and Technology Group at Vanderbilt [CTGV] (1996) has developed a framework for understanding technology and education research. We will refer to that framework as the CTGV framework. The framework is important because, in addition to technology, it includes explicitly two important situational variables discussed above that are central to school reform: the goals and the context for teaching and learning. A distinction is made between transmission models and constructivist models.

Transmission models, also referred to as teacher-centered models, are based on the following assumptions:

- Learning involves the accumulation of particular sets of facts and skills.
- Teaching involves the transmission of facts and skills by an expert.
- Assessment involves an accounting of whether the desired facts and skills have been acquired.

Constructivist models, or student-centered models, are based on assumptions different from the transmission models and are better aligned with the principles of the school reform movement:

- Learning results from the construction of knowledge when the learner interacts with the physical and social environment; since knowledge and experience play an important role in thinking, we should expect individual differences in development and in readiness for various types of learning; students should have opportunities to plan and organize their own learning and problem solving and to work collaboratively with others.
- Teaching involves the creation, management, monitoring, and evaluation of appropriate environments to facilitate learning; the teacher is always learning from his/her students; sustained thinking and opportunities for in-depth explorations such as those available in project-based courses replace

survey courses; communication patterns are richer and much more complex than in a transmission model.

- Assessment goes beyond accounting for factual knowledge, low-level skills and memorization of procedures; it should include "direct measures" and performance-based assessments; it should be formative and designed to encourage reflection and learning (CTGV, 1996).

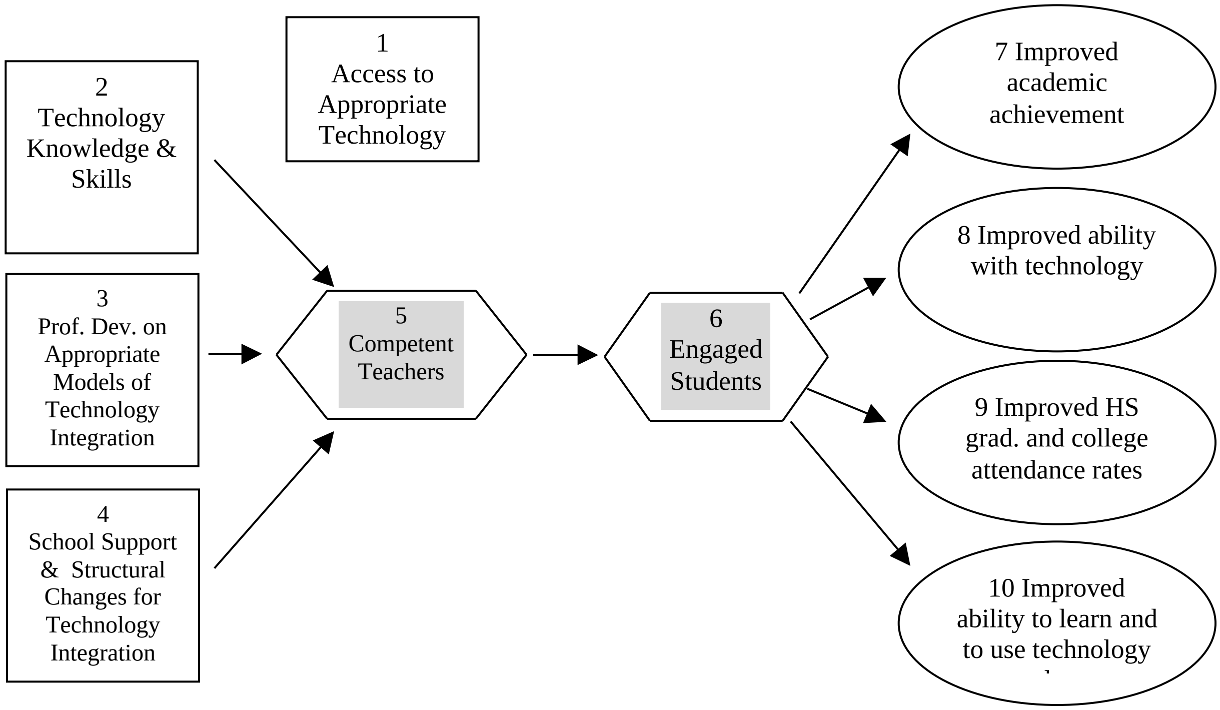
The CTGV framework makes a further distinction between constructivist models involving only *part* of the school day and models that are applicable to *all* schooling activities. With respect to the educational context, the CTGV model makes a distinction between individual classes and schools, and connected (or networked) classes, schools, and communities.

Both the distinction between transmission and constructivist models (involving either part of the school day or all schooling activities) and the distinction between individual and networked classes or schools have direct implications for technology integration. As stated earlier, student-centered models are better aligned with the goals of school reform and more likely to put powerful technology tools in the hands of learners. This is particularly important when learning objectives involve higher-order thinking skills, such as the presentation of information in new forms, problem-solving, the analysis of data, the visual representation of ideas, or the analysis and evaluation of a text. Student-centered models also appear to be more appropriate when assessment includes performance-based assessments and the involvement of learners in the identification of the criteria of successful performance. With respect to networks, the degree of interactivity and access to multiple sources of information are likely to increase with the size of the network.

The Process of Technology Integration

Figure 1 depicts a simple logic model to better understand the *process* of technology integration in schools. The model is an attempt to link inputs (represented by rectangles) to outputs or terminal goals (represented by circles). Hexagons stand for critical intermediate goals. This model is based on the notions of *causal mapping* (Venezky, 1999), *systems thinking* and *task analysis* (Romiszowski, 1986), as these apply to the design, implementation, and evaluation of complex interventions in education and other fields. A similar model was developed for the evaluation of **Alliance+: Professional Development Through Technology**, a TICG project (Friedman, 2000; Yepes-Baraya, 2000b).

Figure 1. Logic Model for Evaluation of Technology Integration



NOTE: 1-4 are inputs; 5-6 intermediate outcomes; 7-10 outputs

The logic model assumes that the process of technology integration as an educational innovation occurs over time and only if the inputs included in the model are present. In that respect the model is prescriptive. Although only four main inputs have been identified, depending on the characteristics of specific projects other inputs might be included. The first input is access to appropriate technology (for learners and teachers). The second input is technology knowledge and skills so that learners and teachers are comfortable with a variety of technology tools for learning and teaching. The third input is professional development on appropriate models for the integration of technology into the learning environment that teachers can use in the classroom. The fourth input corresponds to the support and structural changes required to effectively integrate technology in schools. The *innovations* provided by TICG and SSP schools result from the interaction of these inputs and the participating teachers in the context of their schools. These innovations consist of *competent teachers, with the capacity to integrate technology in school, that are supported by the school's administration and the community to take full advantage of the unique resources that technology can provide.* Competent teachers are considered to be a critical intermediate goal – and not a terminal goal – because their improved teaching ability and facility with information technology are means for achieving specific student goals and not ends in themselves. The same can be said of engaged students, the other critical intermediate goal. Student engagement is a necessary but not a sufficient condition for improved academic achievement and improved ability to use technology.

Examples from Star Schools and Challenge Grant Evaluations

The Construct of Technology Integration

The logic model assumes that technology integration is a process that happens over time and requires certain conditions, referred to as *inputs* in the model, to take root in the classroom. At the core of the construct of technology integration are the two key *intermediate outcomes*: competent teachers and engaged students. The measurement of technology integration can be approached in two steps. The first step involves measuring directly the two key intermediate outcomes (elements 5 and 6 in the model). This would result in information relative to teachers' actual competency in technology integration, and students' experiences with technology integration. A second step involves measuring the conditions leading to technology integration (inputs 1-4). This information would complement that obtained in the first step and might be used to identify corrective action necessary to achieve technology integration.

Criteria and Guiding Questions

The extent to which each element (input or intermediate outcome) in the model is present and the appropriateness of its implementation for a particular TICG or SSP project can be determined by identifying the *key criteria* for each element. The sources of the criteria are varied and include the literature, task analysis, and empirical observations. The key criteria, in turn, can be used to generate *guiding questions* to assess the extent to which the criteria are present. Often these general questions need to be broken down into more specific questions. Answers to both general and specific questions for all the elements in the model should generate enough information to describe a given project and begin to assess the extent and appropriateness of technology integration. In the following sections, for each element in the model, the key criteria are identified first, then general questions are listed, and finally sample questions or items from instruments used in TICG, SSP, and other projects are presented.

1. Access to Appropriate Technology for Learners and Teachers

Two main criteria with respect to technology access are: (1) the availability of technology at school and/or in the classroom, and (2) access for *all* students to technology resources. A third criterion that has been identified in the case of Internet-based work is the availability of technology resources (in this case an Internet connection) at home for teachers. Graham (2000) and others suggest that, given the tight schedule of most teachers during the school day, having an Internet connection at home affords them time after school to locate resources and plan activities for their classes. The questions below should help identify the types of technology available, teacher and learner access, and the appropriateness of the technology. The questions can be formulated for teachers, learners, or technology coordinators.

General Questions:

- What types of technological tools are available to teachers and learners?
- Are these tools appropriate for the teaching and learning goals being pursued?
- Are these tools conveniently located?
- Are there enough tools for everyone involved?

Sample Questions:

The questions below comprise the Internet Connectivity and Access section from a teacher technology survey used in the evaluation of **Alliance+ Project** (Yepes-Baraya, 2000a). A few questions are open-ended, but most require making a selection.

- How many computers do you have in your *classroom*?
- What is the student-to-computer ratio in your *classroom*?
- How many of the computers in your *classroom* are connected to the Internet?
- How long have you had an Internet connection in your *classroom*?
- What type of Internet connection do you have in your *classroom*?
- Are you able to display a computer screen to the entire class with a projector or large screen monitor in your *classroom*?
- How many computers connected to the Internet in other locations in your school are available to use with your students?
- Where are these computers located? (Please circle all that apply)

- Is the number of computers connected to the Internet that you and your students have access to in you school sufficient for your purposes?
- Do you a have a home computer connected to the Internet?
- What percentage of your students have a home computer connected to the Internet?

2. Technology Knowledge and Skills of Teachers and Learners

The main criterion for this element is whether teachers and learners have mastered the specific technology knowledge and skills to use technology effectively and efficiently. The questions below can be formulated for teachers or learners.

General Questions

- What specific technology tools are teachers and learners familiar with?
- How comfortable are teachers and learners with these tools?
- What tools would teachers and learners need to learn to work with to be better equipped to integrate technology in teaching and learning?

Sample Questions

In the Teacher Technology Survey used in the **ACT Now! Sweetwater Union High School** project respondents were asked to self-assess their skill level in a variety of applications/software that may be used in technology integration. Scale: (1) no experience, (2) beginner, (3) intermediate, (4) advanced, (5) mastery (Bober, Harrison, & Lynch, 1997).

- Word processing (e.g., Word)
- Presentation (e.g., PowerPoint)
- Multimedia authoring (e.g., HyperStudio)
- Classroom management (e.g., lesson planning, grading)
- Graphics (e.g., Print Shop)
- Courseware (e.g., Carmen San Diego)
- Web browsers (e.g., Netscape)
- E-mail (e.g., Eudora)
- Other: ___________________

3. Professional Development on Appropriate Models of Technology Integration

The main criteria for measuring and evaluating this component are: (1) the availability of specific curriculum models for teachers to integrate technology in the classroom, and (2) technical support and resources to achieve this end. The literature suggests (e.g., Venezky, 1999) that teachers need specific models to integrate technology in the classroom that go beyond the information provided in a basic technology skills course. Effective models provide information relative not only to curriculum, technology, and assessment practices, but also about changes in the traditional roles of teachers and students. Some general questions that might be asked to elicit information regarding this input refer to the participation in or need for professional development focusing on technology integration. These questions can be asked of technology coordinators or teachers. Other questions pertain to knowledge of or use of specific models of how to integrate technology into the curriculum.

General Questions:

- In what topic areas have teachers received (or been provided) professional development for technology integration?
- How useful/applicable did teachers find these professional development activities?

Sample Questions:

The following questions were adapted from a Teacher Questionnaire for the **Alliance+ Project** (Yepes-Baraya, 2000a). "Please check the topics that you have received professional development on and rate how useful you found these activities." A five-point scale was provided: 1=did not attend, 2=not useful, 3=little useful, 4=useful, 5=very useful.

- Understanding the difference between technology integration and technology use
- Understanding the roles of teachers, students, and technology in collaborative learning
- Teaching and learning in a project-based environment
- Identifying technology resources for standards-based educational objectives for a given subject
- Using instructional design to match educational objectives to technology tools
- Developing technology-based lesson plans
- Developing performance-based assessments

   - Working with students to develop scoring rubrics for performance assessments.

- Approximately how many hours of professional development on technology integration have the majority of teachers in your school received in _________ (a given period of time)?   Five categories are presented, ranging from "0-5 hours" to "more than 20 hours.".

- Approximately what percentage of the following types of teachers in your school has participated in professional development on technology integration in _________ (a given period of time)? (Four categories are presented, ranging from "none" to "over 75%.").

   - Math teachers
   - Science teachers
   - History teachers
   - Social Studies teachers
   - Foreign Language teachers

- Approximately what percentage of the teachers in your school has received professional development on technology integration in _________ (a given period of time)? (Four categories are presented, ranging from "none" to "over 75%.").

4. School Support and Structural Changes for Technology Integration

In addition to access to appropriate technology tools, basic technology skills, and knowledge to use technology as a regular component of teaching, it is commonly accepted that teachers require support from their own schools and communities in order to effectively integrate technology into the curriculum. The biggest challenge for many teachers occurs when they return to their classrooms and attempt to apply in their teaching what they learned in a training workshop. Hargreaves, Moore, and James-Wilson (1997) suggest that teachers need time to talk about their experiences as innovators and "think their way through new practices with colleagues." Venezky (1999) has remarked that few technology-in-education programs explicitly include organizational changes as one of the required inputs for the successful integration of technology in the classroom. In his review of systemic reform, Fullan (1994) argues that neither top-down nor bottom-up strategies for educational reform work. He proposes a sophisticated blend of the two. Hence, the general criteria for the evaluation of this component are administrative support to make time, resources, and support services available to teachers, and the willingness of teachers to integrate technology in the classroom (Yepes-Baraya, 2000).

General Questions:

- What type of support/incentives have been provided to assist teachers in using technology to enhance learning?
- What types of problems/barriers have teachers encountered that prevent them from integrating technology into the curriculum?

Sample Questions:

The following options were included in the Showcase Winners Survey done for the **Iowa Distance Education Alliance**. (Maushak, 1999):

- Release time for planning
- Schedule changes for teachers that allow collaborative planning and learning
- Technology resources for classroom and/or media center
- Technology certification
- Expectation/requirement that faculty use technology as a learning tool
- Acknowledgement/recognition of effective teacher use of technology
- Salary incentives for teachers using technology as a learning tool

The following questions were adapted from a Teacher Questionnaire for the **Eiffel Project** *(*Denes, 1999).

- Please indicate the extent to which the following are barriers to integrating technology into your classroom instruction. (Four categories are presented, ranging from "none" to "a great deal.")
    - lack of time to learn about technologically-enhanced instruction
    - lack of time to develop technologically-enhanced curriculum
    - lack of time in school schedule for projects involving technology
    - lack of technical support for technology projects
    - lack of knowledge about ways to integrate technology to enhance curriculum
    - lack of administrative support for integrating technology into the curriculum

5. Competent Teachers (Teachers who can use appropriate technology to achieve diverse desirable learning goals)

Having teachers who are proficient with technology to further both basic and complex student learning is a critical intermediate goal in the process to achieve effective technology integration. The key criteria to identify teachers who are competent with technology involve: (1) identifying behaviors that correspond to those factors included in the CTVG framework, i.e., the extent to which teaching and learning are teacher-centered or student-centered (and the resulting curriculum, instruction, and assessment practices), and the extent to which a given class is part of a larger network; and (2) identifying behaviors that reflect the frequent and effective use of powerful technology to achieve desired teaching and learning goals. Questions to measure the attainment of this intermediate outcome should include questions that describe the roles of teachers and learners, learner characteristics and the learning environment, curriculum standards and teaching and learning goals, and how frequently and effectively technology is used to achieve these goals.

General Questions

- What proportion of the learning activities are teacher-centered and what proportion are student-centered?
- Is the class part of a larger network of teachers and learners? If so, describe the network.
- What are the characteristics of the learners in the class?
- What are the main teaching and learning objectives of the lesson?
- What specific technology tools do teachers and students use to achieve these objectives?
- How frequently do they use these tools?
- What models or activities do teachers use to with students to integrate technology into the curriculum and how appropriate and useful are they? How is learning being assessed?

Sample Questions

The following questions about teaching practices were adapted from a Follow-Up Teacher Questionnaire for the **Teacher Led Technology Project** (Berkeley Planning

Associates, 1999). These questions are designed to elicit information regarding teacher and student roles and to identify changes resulting from participation in the project.

- In what ways, if any, has participating in the project, including both the new technology in your classroom and participating in professional development activities, changed your instructional practices? As a result of participating in the project, I: (The following four options are available for each of the items below, "yes," "no," "not yet, but I think I will," and "doesn't apply").
  - Spend less time lecturing the whole class.
  - Spend more time with individual students.
  - Am more comfortable with small group activities.
  - More often use cooperative learning techniques.
  - Am more comfortable with students working independently.
  - Am better able to differentiate instruction for students with different abilities or learning styles.
  - Am better able to teach problem solving and/or critical thinking.
  - Am better able to present complex material to my students.

How often do you as a teacher do the following? As a teacher I do the following: (The following five options are available for each of the items below, "never," "less than monthly," "1-4 time per month," "5-10 time per month," and "more than 10 times per month."

- Integrate technology into a lesson in a core subject area.
- Use computer-based activities, materials, or discussions.
- Use TV-computer hookups or other presentation technology for whole class or group lessons.
- Use a computer or other technology to assist a student with a learning problem. Please describe: _________________
- Use a computer or other technology to assist a student with limited English proficiency. Please describe: _________________
- Use a computer or other technology to plan curriculum and lessons.
- Exchange ideas about technology/resources with other teachers in the district or in the school.
- Use technology to produce communications for parents.

The following Guidelines for Elementary/Middle Classroom Observations were adapted from the **Teacher Led Technology Project** (Toms Barker, 1998). The following are the key classroom and teaching and learning strengths that are central to the goals of the project -- the things to look for while you are observing in the classroom:

- Technology integration
  - Software and computer activity links to other activities, discussions, projects in the classroom.
  - Computer and other technology are accessible to all, including those in wheelchairs.
  - Computer and other technology are used in a group setting, as part of a group project.
  - Software, computer activity supports core learning goals in math, language arts, science, etc.
  - Several different skills are being developed at once, in an integrated way.
  - Student products are being generated (if applicable) through the use and influence of the computer (products may include stories, published "books," art work, graphics, graphs, charts, video, research reports).
- Cooperative learning
  - Groups of students work cooperatively at computer or with other technology
  - Students solve problems, reach goals as a group.
- Student initiative/leadership/direction
  - Students talk more than teacher.
  - Student ideas drive projects.
  - Students manage, organize projects.
  - Students help, talk with other students when appropriate.
- Diversity accommodations
  - Students with different characteristics participate equally and in an integrated way with equal access to computer use. The teacher offers varying approaches to student with different needs or learning styles.

   - Teacher divides attention fairly, equally among students.
- Technical issues
   - Teacher and students are proficient with both software and hardware.
   - Students help each other if technical problems arise.
   - The teacher responds effectively to any technical problems or questions from students.
- Overall impression
   - Energy level in the classroom is high.
   - Students are interested, motivated, engaged in learning.

The following questions about computer use were adapted from a Teacher's Survey from the 1998 Teaching, Learning, and Computing study (Becker & Anderson, 1998). “Which of the following are among the objectives you have for student computer use? Check all that apply:”

- Mastering skills just taught
- Remediation of skills not learned well
- Expressing themselves in writing
- Communicating electronically with other people
- Finding out about ideas and information
- Analyzing information
- Presenting information to an audience
- Improving computer skills
- Learning to work collaboratively
- Learning to work independently

6. Engaged Students (How Technology is Used by Learners and for What Purposes)

This is the other critical intermediate goal on the path to improved learning and student outcomes that offers indicators of technology integration. The literature offers many examples on a small scale of how technology can be used to transform teaching and learning (see David, 1994). One factor these experiences with technology have in common is that they are based on the premise that understanding and problem solving require activities that engage students in constructing knowledge. Student engagement can be manifested in different ways, but overall it is likely to include more time spent on task in and out of school, more self-directed learning, and increased participation in special projects.

General Questions:

- What specific technology tools do students use in the classroom? At home?
- How frequently do they use these tools?
- What learning goals are these tools contributing to?
- What cognitive processing is required by the technology used?
- What is the level of commitment of students to technology-based learning?

Sample Questions:

The following questions were adapted from a Follow-Up Teacher Questionnaire for the **Teacher Led Technology Project** (Toms Barker, 1999) "How often do you have your students use technology? I have my students use technology to:" (Scale: never, less than monthly, 1-4 time per month, 5-10 time per month, and more than 10 times per month.)

– Learn from software that teaches academic content.
– Develop written products or graphics/artwork.
– Develop multimedia products (slide-shows, videos, HyperStudio stacks, etc.)
– Analyze or display data (using spreadsheets, graphing software, etc.)
– Do research via CD-ROM.
– Do research via the Internet.
– Use the Internet in other ways (please describe how)
– Review, practice, reinforce basic skills (please specify software)
– Deepen understanding or develop critical thinking skills (please specify software)

The following questions were adapted from a Student Survey for the **Eiffel Project** (Denes, 1999). During the *past week* approximately how much time did you spend on your *home* computer doing the following activities?: (Scale: none, less than 1 hour, Between 1-2 hours, Between 2-3 hours, and More than 3 hours.)

– playing games
– playing educational games
– visiting chat rooms
– typing homework assignments
– "surfing" the Internet
– e-mailing
– researching subjects of interest to you on the Internet
– researching subjects of interest to you on a CD-ROM
– looking for information for school projects on the Internet
– looking for information for school projects on a CD-ROM
– completing projects and portfolios for schools

During the *past week* approximately how much time did you spend in the *computer lab* doing the following activities? The same activities listed in the previous question are listed.

How often do you use computers in the following classes? (Scale: Never, Twice a month, Once a week, A few times a week, and Everyday.)

– English
– Math
– Science
– Social Studies
– Foreign Language/ESL

During the *past week* about how much time did you spend doing the following activities on the computer in each class (same scale)? The list below is for English. Similar lists were provided, tailored to math, science, social studies, and foreign languages.

– typing written text
– composing text on a computer
– researching for an assignment
– working on a project
– developing a Web site
– working on your portfolio
– creating presentations

7. Accountability and Long-Term Goals

The measurement of these components of the model goes beyond the measurement of technology integration. It is important, however, to ascertain the extent to which technology use has contributed to the improvement of teaching and learning. With the emphasis of the TICG and SSP projects on reaching *all* students, measuring access and achievement of under-represented minority students relative to the majority students should be one of the main evaluation criteria. Achievement should be defined broadly to include not only achievement in the key curriculum areas into which technology has been integrated (e.g., science, mathematics, language arts, social studies, or foreign languages), but also improved ability to use technology for different purposes, improved attendance and improved graduation rates, and improved ability to learn and to use technology for this purpose. The reader is referred to other chapters in this sourcebook for measures for some of these long-term goals.

Discussion

The model of technology integration presented in this chapter assumes that technology integration is a process that happens over time and requires certain conditions and resources, referred to as *inputs* in the model, to take root in the classroom. Hence the general and sample questions provided for each element in the model can be used to obtain *formative* and *diagnostic* evaluation information about a given project.

If instead the evaluation goal were to obtain *summative* evaluation information on a well established project, it might be more appropriate to concentrate on questions regarding the achievement of the *intermediate outcomes* identified in the model, namely, competent teachers and engaged students. These are questions at the core of the concept of technology integration, namely, questions focusing on new forms of classroom organization and new roles for teachers and learners, the existence of computer networks, teacher and learner use of a variety of technology tools, how frequently these tools are used, the specific skills and knowledge targeted with the technology, and the appropriateness of the technology to achieve the desired teaching and learning goals.

Answers to each of the above questions provide information that combined with answers to the other questions can be used to obtain a broad measure of technology integration in a class, school, or school district. Let's assume, for example, that for a given eighth grade science class evidence has been collected about the following: throughout the year there is a balance between teacher-centered and student-centered activities; the class is part of a larger network involving other middle school science classes in schools across the country and scientists at a local hospital with whom the students correspond biweekly about project-based work involving authentic investigations; students work in teams in the lab and use probeware and other electronic equipment as part of their project; students use PowerPoint to make monthly presentations of their findings that are part of a performance assessment jointly designed with the teacher; under the teacher's supervision, the more advanced students often help those who may not be as well prepared. On the other hand, let's assume that in the same town evidence collected on a tenth grade English class shows the following: the great majority of class activities are teacher-centered; the teacher's main goal is to help students improve their writing skills; students are being trained to follow a theory-based model about the composing process; although there is one computer in the classroom connected to the Internet, the teacher is the only person who uses it, mostly to consult the Encyclopedia Britannica, the Oxford English Dictionary, and other online reference services; the teacher also uses the classroom computer to e-mail other teachers about selecting composition topics for his students; every Wednesday and Friday students spend one hour in a non-networked lab where they use Word to copy the compositions they work on at home. A comparison of these two classes suggests that technology has been more fully integrated in the science class than in the English class. Further, in the science class, technology is being used not only to attain simultaneously several real-world, complex skills, but also to take advantage of the interactivity and access to multiple sources of information that computer networks and electronic mail can provide. In conclusion, both the technology itself and the how it is being used are better aligned with OERI's vision of educational technology in the science class than in the English class.

The approach to the measurement of technology integration discussed in this chapter requires the measurement of different aspects of technology integration, some of which go beyond technology and have to do with the purposes and the context of technology use. This approach relies on using various scales to measure quantitative aspects of technology integration, such as frequency of use of technology, time spent with technology, proportion of activities involving technology, network size, student/computer ratio, proportion of activities that are student centered, etc., and judging not only the degree of integration of technology in the curriculum but also the appropriateness of the technology to achieve teaching and learning goals. It would seem desirable to combine these measures into an overall measure of technology integration. This goal, however, is not so easily accomplished because not all aspects should be weighted the same and there is still no generally accepted formula. Attempts to simplify the measurement of technology integration often result in rough measures, like the self-assessment rubric presented below (Bellingham Public Schools, 2000), that assume a shared understanding of this domain.

> ___Level 1 - I do not blend the use of computer-based technologies into my classroom learning activities.
>
> ___Level 2 - I understand the district technology plan supports integration of technology into classroom activities, but I am still learning about what strategies will work and how to do it. I accept student work produced electronically, but do not require it.
>
> ___Level 3 - From time to time, I encourage my students to employ computer-based technologies to support the communicating, data analysis and problem solving outlined in the district technology plan.
>
> ___Level 4 - I frequently model and teach my students to employ computer-based technologies for communication, data analysis, and problem-solving as outlined in the district technology plan.

Accordingly, one challenge that researchers and evaluators will face in the near future is to search for better, more holistic definitions of this rather elusive domain. This challenge will be compounded by the need to account for characteristics of emerging technology, such as miniaturization, portability, and reduced cost (Means, 2000), that will result in technology that is more user-friendly and accessible, and that provides opportunities for greater interactivity. .

References

Becker, H.J., & Anderson, R.E. (1998). *Teacher's survey: Combined versions 1-4.* (Teaching, Learning, and Computing): 1998. Irvine, CA: U.C. Irvine.

Bellingham Public Schools (2000). *Staff Use of Technology – 1999-2000 Self-Evaluation Rubrics.* Bellingham, WA: Available online: http://www.bham.wednet.edu/tcomp.htm

Bober, M., Harrison, P.J., & Lynch, E.W. (1997). *Teacher Technology Survey.* (ACT Now! Technology Challenge Grant). San Diego, CA: San Diego State University.

Cognition and Technology Group at Vanderbilt. (1996). Looking at technology in context: A framework for understanding technology and education research. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of Educational Psychology* (pp. 807-840). New York: Simon & Schuster MacMillan.

David, J.L. (1994). Realizing the promise of technology: The need for systemic reform education. In R.J. Anson (Ed.), *Systemic reform: Perspectives on personalizing education*. U.S. Department of Education, Office of Educational Research and Improvement. http://www.ed.gov/pubs/EdReformStudies/SysReforms/

Denes, V. (1999). *Eiffel Project student, teacher, and technology coordinator questionnaires.* New York: Columbia University, Teachers College, NCREST.

Expert Panel on Educational Technology (1999). *Guidelines and Materials for Submitting Educational Technology Programs for Review*. Washington, DC: U.S. Department of Education. Available online: http://www.ed.gov/offices/OERI/ORAD/LTD/final/app99pkg.html

Friedman, E.A. (2000). *Conceptual framework and organizational structure of Alliance+: A national model for internet-in education professional development.* Hoboken, NJ: Stevens Institute for Technology, Center for Improved Engineering and Science Education.

Fullan, M.G. (1994, September). Coordinating top-down and bottom-up strategies for educational reform. In R.J. Anson (Ed.), Systemic Reform: Perspectives on Personalizing Education. U.S. Department of Education, Office of Educational Research and Improvement. Available online: http://www.ed.gov/pubs/EdReformStudies/SysReforms/

Graham, K. (2000). *From technologically challenged to savvy cyber teacher: Lessons learned from participating in the Alliance+ project.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Hargreaves, A., Moore, S., & James-Wilson, S. (1997). How teachers deal with secondary school change. In *Research in Ontario Secondary Schools, 3*. University of Toronto, Ontario Institute for Studies in Education. Available online: http://www.oise.utoronto.ca/~fieldcen/vol3no3.htm

Jonassen, D. H., & Reeves, T.C. (1996). Learning with technology: Using computers as cognitive tools. In D.H. Jonassen, *Handbook of research for educational communications and technology*. New York: Simon & Schuster MacMillan.

Maushak, N. (1999). *Showcase winners survey.* (Iowa Distance Education Alliance.) Johnston, IA: Iowa Public Television.

McNabb, M., Hawkes, M., & Rouk, U. (1999). *Critical issues in evaluating the effectiveness of technology.* Paper presented at the Secretary's Conference on Educational Technology: Evaluating the Effectiveness of Technology, Washington, DC. Available online: http://www.ed.gov/Technology/TechConf/1999/confsum.html

Means, B. (2000). Technology use in tomorrow schools. *Educational Leadership, 58*(4), 57-61.

Romiszowski, A. J. (1986). *Designing instructional systems – Decision making in course planning and curriculum design.* New York: Nichols Publishing.

Toms Barker, L. (1998). *Guidelines for elementary/middle classroom observations.* (TLTC Project.) Oakland, CA: Berkeley Planning Associates.

Toms Barker, L. (1999). *Follow-up teacher questionnaire.* (TLTC Project.) Oakland, CA: Berkeley Planning Associates.

Venezky, R. L. (1999). *A guide to causal mapping*. Paper presented at the OERI Technology Evaluation Institute, Ann Arbor, MI.

Yepes-Baraya, M. (2000a). *Teacher Technology Survey.* (Alliance+.) Princeton, NJ: Educational Testing Service.

Yepes-Baraya, M. (2000b). *Lessons learned from the evaluation of Alliance +: An internet-in-education professional development program*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

7. The Evaluation of Dissemination Efforts

Paula Szulc Dominguez
Hezel Associates

Introduction

Information sharing is at the heart of evaluators' activities. Our roles as evaluators of technology initiatives funded through the U.S. Department of Education's Office of Education Research and Improvement (OERI) call on us to respond to a complex set of reporting tasks. We document our findings through written and oral reports to project staff, who may, in turn, pass on the information to stakeholder groups and project participants. We also assess the extent to which our respective projects share their work with stakeholder groups through dissemination activities, and submit these results to OERI in our annual reports. We are professionally dedicated to the larger issue of increasing our discipline's understanding of dissemination practices through our findings and observations, and describe our work at the conferences and meetings we attend.

Two factors encourage the belief that examining dissemination activities within the context of project evaluation is a priority issue. On the one hand, information dissemination itself occupies a large proportion of evaluators' time. On the other hand, educational initiatives (especially those that are publicly funded) are honor bound to share their results, with the goal of further improving students' lives. Many also include dissemination as a project task or intended outcome as part of the overall project design.

In fact, however, evaluators of educational projects have devoted minimal attention to understanding how well project dissemination activities take place. This lapse means that the underlying purpose of project evaluation—to ensure that members of a decision- making community become aware of findings (Shrinkfield, 1983)—remains largely unexamined.

To what extent does the appropriate information fall into the right hands? To what extent is the information accurate? More importantly, how were the data gathered, and to what extent is the information useful to its intended audience? Those of us who are charged with evaluating educational technology initiatives have few resources to call upon when we consider the effectiveness of dissemination efforts and questions such as these. This sourcebook chapter represents an initial effort to scrutinize dissemination activities and provide authentic examples of how educational technology and other evaluations have gone about assessing dissemination. It begins by identifying some of the issues involved with evaluating dissemination outcomes using findings from other disciplines. It then suggests a framework for examining the merit of dissemination activities. Finally, it presents relevant examples from educational technology and other evaluations to illustrate key points.

Throughout this chapter, the word "technology " is used frequently. Although technology typically refers to equipment or products when it appears in discussions among educational technology researchers, this discussion relies on its broader definition as an idea, practice, or product that is used as a tool to facilitate human activities. Thus, findings resulting from educational technology research are considered a form of technology, in the same manner as is a CD-ROM.

What is Dissemination?

Dissemination refers to the diffusion and adoption of ideas, practices, programs, and products by particular audiences, including individuals, agencies, and institutions (Glik, Berkanovic, Macpherson, Strone, Ratner, & Connell Jones, 2000). The fundamental ideas underlying dissemination stem from the field of communication research, more specifically from work that centers on the diffusion of innovations. From its inception, the concept of diffusion closely intersected with technology; a study of the diffusion of a technological innovation (e.g., a farmers' adoption of hybrid corn, Ryan & Gross, 1943) is considered to be the defining example of early diffusion research in the United States.

Essential to diffusion is the notion that the idea, practice, or product that is communicated is new. This *innovation* is one of the four basic elements of diffusion (Rogers, 1995). A second component of diffusion is *communication*, the particular exchange of information about the new idea by means of a communication channel. The communication channel may be a form of mass media (e.g., television and radio spots) or interpersonal, face-to-face contact (e.g., groups of colleagues, members of clubs). The communication of the innovation takes place over *time*, the third element of the diffusion process. The time dimension describes several aspects of diffusion: (1) the process by which an individual passes from initial knowledge of an innovation to the stages of persuasion, decision, implementation, and confirmation, (2) the relative earliness/lateness that an individual (or group) chooses to adopt an innovation, compared with others, and (3) an innovation's rate of adoption in a system (Rogers, 1995, p. 20). All of this activity takes place in a *social system*, the final component of diffusion. Social systems create parameters that affect diffusion, through norms of communication and behavior, through the roles of opinion leaders and change agents, and through the types of decision making that exist (whether an individual, group, or authority can choose to adopt an innovation).

Over the years, diffusion researchers have identified factors that can help or hinder the sharing of information about new technologies. Four factors in particular have direct implications for the information and products generated and disseminated by educational technology initiatives:

- the relative advantage an innovation offers;
- compatibility with existing technologies;
- the characteristics of change agents and
- the size and structure of organizations.

Relative Advantage

A technology's relative advantage refers to the degree to which it is perceived to offer enhanced benefits compared to those of existing technologies. Those benefits can be expressed in terms of economic advantages—when a new product (e.g., an electronic field trip to a national park) can be produced and sold less expensively than an old product (the traditional class field trip). Alternatively, the economic advantage of a new technology may assume the appearance of increased efficiency, such as in a busy adult learner's ability to complete a degree through distance learning. The advantage can simply be one of availability, such as in the case where a school is located too far away from a given location for a field trip to be considered feasible. The social prestige a new technology offers can also lend it an advantage. For example, a teacher who participates in online workshops may be attracted by the idea of being seen among her peers as a person who is open to new approaches to professional development. Incentives tied to the use of the innovation can also create a relative advantage. For example, faculty members who integrate multimedia technology into their courses may receive special stipends to compensate them for their work.

Compatibility

How closely a new technology aligns with existing ones will influence its adoption—the more compatible the new technology is, the more likely it is to be adopted. Compatibility can be expressed in terms of existing values. For example, technology products that build on national content standards align well with accepted beliefs among educators, and are therefore more likely to be incorporated into the curriculum. Another facet of compatibility involves a harmony with previously introduced ideas. Educational technology research that builds on established methods and theoretical perspectives is more apt to be attended to than research that makes no reference to the established literature. Compatibility with prior experiences is also important. That is, an online workshop for teachers will tend to be more successful if it includes the array of information, materials, and activities teachers usually encounter in traditional, face-to-face workshops.

Characteristics of Change Agents

A change agent is a person who assumes responsibility for encouraging the adoption of a technology within a particular audience. Typically, the change agent is not directly affiliated with the experts who originated the technology and is familiar with the needs of the intended clients. The change agent therefore acts as an important intermediary between the originators of a technology and the intended user group. A change agent's identity and credibility, his or her understanding of client needs, and the communication efforts that takes place between change agent and client all impact the rate of technology adoption.

Size and Structure of an Organization

The bottom line in any education initiative is student learning. In order to reach students, education initiatives must navigate through schools and school districts that differ in size, complexity, and character. It is therefore important to keep in mind the particular characteristics of organizations that contribute to or frustrate the adoption of new technologies. In general, larger organizations with decentralized power structures tend to be more innovative than their smaller, more centralized counterparts. That is, large, decentralized organizations often have a range of resources available and a network of leaders who can introduce new ways of conducting business at multiple levels, although system-wide change may be more difficult in larger decentralized organizations than in smaller centralized ones.

An important consideration for evaluators interested in dissemination is the notion that the adoption and implementation of technologies within organizations is not a "yes/no" situation. In fact, it is the exception rather than the rule that an innovation will be embraced whole-heartedly. Instead, organizations like schools and school districts will opt to reinvent a technology by adapting it to local needs. In this manner, certain, locally palatable components of the technology are accepted, while others that do not align with local tastes are ignored. Individuals involved in the dissemination of educational technology ideas and products can expect to find a range of adoption practices, rather than a dichotomous array of adopters vs. non-adopters.

Diffusion research, then, provides a launching point for understanding two aspects of the dissemination of educational technologies:

- how to structure and modify ongoing dissemination activities, and
- how to examine the effectiveness of dissemination efforts.

As evaluators, our examination of dissemination efforts needs to distinguish between and assess (1) the activities that are conducted and (2) the impact they realize. Before we turn our attention to discussing how to evaluate dissemination efforts, let us first consider the range of dissemination activities that typically characterize educational technology initiatives.

Dissemination Activities

As we've described above, dissemination efforts involve the communication of a new idea or product over a period of time through a particular social system. Within educational initiatives (as well as any other), the communication that takes place can be less than what project directors had hoped for and intended. Information can be misconstrued by target audiences that do not share the same language or approach as those involved in the project. Alternatively, accurate information can be distributed, but to the wrong people. Perhaps the worst case scenario is dissemination efforts whose results are summarized in thick, bound reports that lie neglected and unused on desks and shelves.

The kind of outcomes associated with dissemination are largely fashioned during an initiative's formative stages. When submitting a proposal, project administrators usually list the kinds of dissemination activities that will take place at the end of the project or, less frequently, over the course of the initiative. Seldom, if ever, do proposals describe whether and how dissemination efforts will be critiqued, perhaps because such scrutiny rarely takes place.

But the actual information sharing activities are (or should be) an outcome of a series of prior decisions about dissemination, and it is these decisions that tend to be overlooked by project administrators and funding organizations alike. Dissemination activities should spring from a plan that articulates the decision making that occurs as the project unfolds. Ideally, an education initiative will devise a dissemination plan during a project's initial phase, with some components solidified at the proposal stage. The first step involved with putting together a dissemination plan requires that target audiences be identified (this step can be developed in the project proposal stage). For each audience, the goals for dissemination (e.g., changes in awareness, changes in knowledge, changes in behavior) are then described. Note that the particular goals associated with dissemination may differ to some extent from those of the project as a whole. Project personnel may next identify in their plans organizations that can act as potential intermediaries (change agents) during dissemination. Only then are actual dissemination activities and communication channels selected. Dissemination milestones and measures can then be identified and incorporated into the project's ongoing administration. Finally, the outcomes of the dissemination efforts can be assessed in light of each stage of decision making.

Table 1: Typical dissemination activities

Audience/ Stakeholder Group	**Dissemination Activity**	**Format**
Policy makers	Policy briefs	Text, electronic
Researchers and Practitioners	Research reports	Text, electronic
	Refereed articles	Text, electronic
	Professional journals	Text, electronic
	Conference presentations	Text, face-to-face
Adult learners/students	Textbooks	Text
	Training manuals	Text, electronic
	Demonstrations	Face-to-face
	Educational materials	Text, video, audio, electronic
Multiple audiences	Web site	Electronic
	Email listserv	Electronic
	Articles in popular press	Text
	Workshops and seminars	Face-to-face

The nature of the dissemination activities, then, reflects the particular audiences or stakeholder groups, dissemination goals, and available resources and relationships. As projects that tend to reach out to multiple target audiences, educational technology initiatives rely on a number of dissemination activities to share information (see Table 1).

It is a straightforward task to describe dissemination activities, since the products are usually tangible. The number of students who participated, the number of CD-ROMs that were distributed, the titles of articles that were published, the content of the workshops that were delivered—all of these can be easily summarized in a final report. Measuring the effectiveness of each activity, however, is a more difficult undertaking.

Evaluating the Impact of Dissemination

As mentioned above, our analysis of project dissemination efforts should include a consideration of how activities unfold (process) as well as the impact that is achieved (outcomes). If a project develops a dissemination plan, then process and outcome evaluations are much easier to conduct—the measures and indicators for effectiveness can be developed readily from the clearly identified target audiences, goals, and communication channels. If a project has not created a dissemination plan, makes its decisions on an ad hoc basis, and leaves the evaluation of dissemination unexamined until the close of business, then the evaluators lack a blueprint for understanding why particular decisions were made and with what intended effects.

Including an examination of how dissemination activities unfold reflects the belief that dissemination should be a dialog between the originator of the technology and its intended audience. That is, instead of structuring the dissemination activities as a one-way flow of information, more sophisticated dissemination efforts include mechanisms to ensure that feedback from target audiences is obtained on a regular basis and incorporated into the activities. The soliciting, examining, and reporting of feedback on dissemination activities is tailor made for each project, and an analysis of dissemination efforts can be incorporated into other formative evaluation activities.

A project's overall success in disseminating a technology as planned takes place at the project cycle's conclusion. Practically speaking, such an analysis is often difficult to conduct because many education initiatives do not generate findings or products until the project's very final stages. Limited time or funding is available to consider dissemination outcomes, a situation that underscores the need for effective planning at the project's outset and for ongoing data collection.

But what should evaluators actually measure in their examination of the process and outcomes of dissemination? At a minimum, evaluators should focus on four areas in order to assess the effectiveness of dissemination: audiences, communication, consequences, and improvement. Each area flows directly from the project's dissemination strategy, and each can be assessed through process and outcome evaluations.

Audience/Stakeholder Group

It is not enough that information is distributed. The key question is whether the project's *intended* audience is reached. For example, Glik et al. (2000) describe how only 34 percent of a widely disseminated middle school immunization curriculum actually reached the most likely users (teachers and school nurses). A slightly larger proportion was found to be held in limbo in a central office or by non-teaching staff, while a slightly smaller proportion was unused, warehoused, or lost. If evaluators simply looked at the sheer number of curricula that were distributed—close to 3000 copies—then they might consider the project dissemination "successful." It is only when the identity of the recipients is factored in that the actual success can be ascertained. The same criticism applies to educational technology initiatives that produce a web site and simply report "hits" to describe dissemination—unless the identities and characteristics of the web site visitors can be established, the numbers provide limited evidence of dissemination accomplishments.

Communication

How effectively was information presented to a project's intended audiences? This question asks us to evaluate two aspects of communication (Saywell & Cotton, 1999). First, how accessible was the information? Did the information assume a physical form that was readily available to target audience members? If the intended audience is adult migrant workers, then producing web-based resources is of limited value. A second aspect related to communication is the comprehensibility and usefulness of the content. Is the information understood by members of the target audience? Is it relevant to their needs? The media and messages employed in dissemination can be examined throughout a project, and adapted in light of feedback from audience members.

Consequences

What happens as a result of dissemination? What knowledge was gained, behaviors changed, or performance improved? It is always difficult to establish cause and effect in research, especially in education settings where so many competing factors each exert an influence on students' and teachers' learning. Nevertheless, there are ways that we can evaluate the outcomes of dissemination. We can assess target audience members' awareness and knowledge of a technology through traditional surveys. We can also track their use of a technology in observation and follow up studies. Another approach involves taking advantage of the tendency among educators to "reinvent" an innovation. Educators will tend to implement selectively certain components of a technology, and it is this selectivity that can be studied in greater detail. Which aspects of the technology were implemented and which were not? What individual and organizational characteristics contributed to the reinvention? Understanding why educators choose to use or not use different components of a technology can help projects further refine their dissemination strategy.

Table 2: Evaluating dissemination process and outcomes

Evaluation Area	Question	Measure	Method
Audience/ Stakeholder Group	Did target audience members receive the information?	Reach Recipients' identity	Distribution of technology Follow up contact
Communication	How accessible was the information to target audience members?	Alignment between format of technology and audiences' resources	Questionnaire to intermediary organizations
	Was the information understood by target audience members?	Comprehension	Beta testing of messages
Consequences	To what extent are target audience members aware of the technology?	Identification of different components of technology	Questionnaire, interviews
	Did target audience members use the technology? How?	Use	Observation, questionnaire
	Which aspects of the technology were used?	Use	Observation, questionnaire
Improvement	Which organizational factors affected dissemination?	Identification of factors that facilitate and frustrate	Interviews with project personnel
	Which practical factors affected dissemination?	Identification of factors	Interviews with intermediaries
	Which perceptual factors affected dissemination?	Identification of factors	Interviews with target audience members

Improving Dissemination

As evaluators, one of our underlying goals is to contribute to continual improvements in educational initiatives through our research and findings. Our efforts to understand dissemination should also include a focus on uncovering factors that affect the success of dissemination efforts. There are at least three dimensions of effective dissemination that we can focus on. First, we can ask what kind of *organizational* characteristics exist that contribute to or impede dissemination. The characteristics may exist at the level of the originating organization (the project itself), the recipient's organization (schools other education settings), and even the level of the grant program funding the initiative. Another dimension to consider is the practical features that exert an influence on dissemination. What constrains an education initiative's or individual's ability to disseminate a technology (e.g., lack of time, lack of incentives), and how can these constraints be relieved? Alternatively, what practical factors readily permit dissemination (e.g., access to support staff, increased conference funding)? Finally, the perceptions of target audiences might affect dissemination efforts and should be considered. How relevant is the research or product to an audience member's needs?

Examples from Star Schools and Challenge Grant Projects

This section offers three very different examples of the evaluation of dissemination efforts. The first is a Technology Innovation Challenge Grant which funded specific dissemination projects through an RFP process. The second is the Distance Learning Resource Network which is responsible for disseminating information about all of the Star Schools grants. The third is a national dissemination center which disseminates the findings of disability research projects funded by the National Institute on Disability and Rehabilitation Research while testing the effectiveness of new strategies and approaches in achieving dissemination and utilization outcomes.

The Triton Project

The San Diego Unified School District's **Triton Project**, a recipient of a 1995 five-year TICG award, sought to improve student performance in mathematics and science using a combination of technologies in an active learning curriculum with an ocean exploration theme. San Diego public schools were networked with each other and with several area museums, research centers, and Aquariums. One major focus of the initiative was the development and dissemination of curriculum-based projects using the

WebQuest approach (Dodge, 1997), an inquiry-oriented activity that relies on the web for content and focuses attention on students' higher order thinking skills (analysis, synthesis, and evaluation). The WebQuest model presents educators with a series of six steps for building curricula that "scaffold" learners as they explore a topic at progressively deeper, more meaningful levels.

In the fourth year of the grant, Triton project administrators released a call for proposals among participating schools for dissemination efforts that would reflect the project's original four goals, serve as evidence of Triton's sustainability, and broaden the initiative's impact. In an example of authentic modeling, the proposal for Triton dissemination projects relied on WebQuest's multi-step approach (San Diego City Schools, 2000). Five school proposals and two proposals submitted by individuals were successfully funded in March 2000 for a total of about $130,000.

In general, the dissemination activities of successful Triton participants fell into one of three areas: increasing parental and community involvement, increasing school-wide dissemination, and enhancing subject area development. For example, Proyecto Busqueda, a project of Memorial Academy (a junior high school), has concentrated on involving a greater number of teachers to create WebQuests in Language Arts, Spanish, Bilingual or Sheltered Instruction (note that one of Triton's original stipulations was that at least twenty percent of teachers at a school had to participate in technology training). As part of their participation, Memorial Academy teachers work with mentors, implement the curriculum in their own classroom, and present their WebQuest units to other staff members.

The Triton Project illustrates two good lessons with regard to dissemination. First, it is possible to consider dissemination and allocate funding for such activities in the early stages of a grant. Triton project administrators opted to decentralize dissemination efforts through their mini-grant program, which suggests a second good lesson. Because of the different approaches to dissemination taken by schools and individuals, the impact of such activities can be assessed on multiple fronts. Measures of the Triton Project's dissemination efforts can include the number of proposals received, the number of curricular units produced, the proportion of teachers trained or parents involved, or the number of community members reached. Since many of the WebQuest curriculum units are located on the web, the number of visitors (and other web traffic data) can also be documented. As many of the Triton Project's dissemination projects are still under development, it remains to be seen to what extent data about reach, use, etc. are shared.

The Distance Learning Resource Network

The **Distance Learning Resource Network** (DLRN) is responsible for disseminating information about all Star Schools initiatives. DLRN relies on a number of communication channels to share information about K-12 distance learning activities, including a web page, electronic newsletter, conference presentations, online events, and an 800 telephone help line. DLRN responds to the needs of several stakeholder groups, including Star Schools Project Directors, K-12 teachers and administrators, and institutions of higher education. Much of Hezel Associates' evaluation activities to date have centered on three aspects of DLRN's dissemination activities—identifying the people who consult DLRN's web pages, assessing the appropriateness of the content and design of DLRN's new web site, and understanding the consequences of target audience members' use of DLRN's resources. The focus here is on the evaluation activities that have conducted to address the first two aspects of DLRN's dissemination—audience identity and content appropriateness.

As described above, DLRN reaches its target audience members in a number of ways, but its web-based activities reach a much larger group than its other efforts. One difficulty inherent in conducting dissemination activities over the web is the web's characteristic anonymity. A basic challenge facing DLRN (and any other entity which relies on the web to share information) is establishing who actually consults its online materials. Given the anonymity provided by the web, how can DLRN determine how well it is reaching its intended audience? To address this question, we have taken initial steps to try to identify the individuals who visit DLRN's web site. At this point, we are relying on web traffic reports from a software program DLRN employs, *Web Funnel 4*, that categorizes visitors in terms of their email addresses--.edu, .gov, .mil, etc. Although the data provide a baseline understanding of the identities of the visitors, they are still limited in one important way. We have observed that the largest proportion of visitors to DLRN's web site (about 30 percent) use personal email (e.g., Yahoo, Hotmail) rather than institutional accounts. For this reason, DLRN may actually reach a much larger audience of K-12 educators (its primary target audience) than it first appears to, as many teachers, faculty members, and administrators may work from home rather than from schools or district offices. To clarify this, we will contact those visitors who use personal email accounts to ask them their institutional affiliation.

Another important question that disseminators of information such as DLRN must ask is: how do recipients of the information perceive its usefulness? To address this, Hezel Associates conducted an email survey of subscribers to DLRN's online newsletter, *Teaching Through Technology*, to obtain their feedback about the quality and usefulness of the newsletter content. Using a 5 point scale where 1 indicated "poor" and 5 meant "excellent," we asked subscribers to rate the overall quality of the content of the DLRN online newsletter. In addition, we used another 5 point scale where 1 stood for "not useful at all" and 5 indicated "very useful" and asked subscribers how useful to their decision making the DLRN online newsletter content had been. Note that we also posed similar questions to people who registered for DLRN's online course and to volunteers who reviewed DLRN's redesigned web site. In this manner, we hoped to provide DLRN with information that would allow it to refine the appropriateness of the content it delivers to its target audiences, especially among K-12 educators.

Although the individuals who responded to the email survey we distributed gave high ratings to the usefulness of DLRN's web site and newsletter, we experienced considerable difficulty obtaining responses to our email based survey, so we are still considering the extent of the web site's usefulness in our work. We surmised that the lack of response we experienced was due to the length of time that had elapsed between initial contact with DLRN and the survey. For this reason, our future evaluations will target individuals who have recently (i.e., within the past three months) contacted DLRN.

National Center for the Dissemination of Disability Research

Although not a technology initiative and not a recipient of Star Schools or Technology Innovation Challenge Grant funds, the National Center for the Dissemination of Disability Research (NCDDR) has designed a series of strategies for assessing dissemination that cut across disciplinary boundaries. NCDDR urges project staff to go beyond the "bean counting" that typically takes place and examine the project's impact on intended audiences. To this end, the NCDDR (1997) has designed a Dissemination Self Inventory that contains a series of questions that are meant to provoke project staff to think more broadly about analyzing dissemination activities. In my opinion, the NCDDR's Dissemination Self-Inventory is an excellent resource for TICG and Star Schools evaluators to draw upon, and for this reason I include extensive excerpts from it.[1]

The NCDDR urges projects to consider "utilization"—the use of information that is disseminated—as the underlying goal of dissemination strategies. According to

[1] Reproduced with the permission of NCDDR. For a complete copy of the Dissemination Self-Inventory, contact the NCDDR at 1-800-266-1832, or write to NCDDR, Southwest Educational Development Laboratory, 211 East Seventh Street, Suite 400, Austin, TX, 78701-3281. The NCDDR's web site address is http://www.ncddr.org.

NCDDR, five factors influence utilization, and are therefore the categories that organize the 54 indicators within the inventory:

1. User Group—user group(s) or potential users of the information or product to be disseminated;

2. Information Source—your project/organization as an information source, that is, the agency, organization, or individual responsible for creating new knowledge or products, and/or for conducting dissemination activities;

3. Content of Message—message content that is disseminated, such as the new knowledge or product itself, and any supporting information or materials;

4. Medium of the Message—the ways in which the knowledge or product is described, "packaged," and transmitted; and

5. Contextual Consideration for Implementation—content for use of the message, that is, the environmental, personal, and other support needed to use the information or product.

Each of the five categories contains a series of questions that ask project staff to reflect on their knowledge of the activities they have conducted and information they have collected. Each question has a 6-point scale associated with it, where 0 represents "don't know," 1 indicates "no," and 5 means "yes." For some questions, a rating of 3 indicates "sometimes" or "partially." At the end of each category, project staff are asked to total their scores for that section by simply adding up their circled responses. Project staff then divide the total score by a defined number (40, 50, or 70, depending on the category) and then multiply this figure by 100 to generate a percentage. The percentage for that category can then be graphed on a chart provided with the inventory in order to map the project's relative strengths and weaknesses.

User Group

The ten questions contained in the User Group category all reflect an interest in assessing the extent to which the target audience members and their needs were involved in the project's development. For example:

– Does your research design clearly define the intended groups of "users" or beneficiaries of your project's results? (Scale: 0: Don't Know; 1: No; 2 [blank], 3: Partially; 4 [blank]; 5: Yes)

– Was your project's original proposal developed in collaboration with your current intended user/beneficiary groups? (Scale: 0: Don't Know; 1: No; 2-4 [blank], 5: Yes)

– Do you have needs assessment data that identifies the extent of motivation or readiness of your intended user groups to use information in the project's topical area? (Scale: 0: Don't Know; 1: No; 2 [blank], 3: Somewhat; 4 [blank]; 5: Yes)

Information Source

The 14 questions that compose the Information Source category revolve around the interplay among the identities of the project, project staff, collaborators, and user groups. For example: (All use the scale 0: Don't Know; 1: No; 2-4 [blank], 5: Yes)

– Is the project perceived by user groups as an active information dissemination resource?

– Is your project collaborating with other organizations that are equally or more highly visible or interactive with your project's intended user groups?

– Do users frequently (three or more times each work week) contact the project for information in your topical area?

Content of Message

This category of the inventory contains 10 questions that query project staff about the information that is being conveyed and the extent to which user groups were involved in developing the content. For example:

– Is the reading/comprehension level required to understand your project's information analyzed and matched to the characteristics of your intended user groups? (Scale 0: Don't Know; 1: No; 2-4 [blank], 5: Yes)

– Have your user groups been involved in developing content and in field-testing (or other review) and revision of your project information? (Scale: 0: Don't Know; 1: No; 2 [blank], 3: Sometimes; 4 [blank]; 5: Yes)

– Do you share and allow requests for information through multiple means, for example, telephone, fax, mail, e-mail, and other modes upon request? (Scale: 0: Don't Know; 1: No; 2 [blank], 3: Sometimes; 4 [blank]; 5: Yes)

Medium of the Message

The Medium of the Message category poses eight questions to project staff, and asks staff to consider the mix of the delivery channels through which information is made available. For example:

- Does your dissemination strategy include opportunities for person-to-person contact with users? (Scale: 0: Don't Know; 1: No; 2 [blank], 3: Sometimes; 4 [blank]; 5: Yes)
- Are you providing information to users through channels (visual, auditory, etc.) they are known to prefer? (Scale 0: Don't Know; 1: No; 2-4 [blank], 5: Yes)
- Is your project information delivered through existing (not developed by grantee) networks, communication channels, associations/organizations, meetings/conferences, and other venues? (Scale 0: Don't Know; 1: No; 2-4 [blank], 5: Yes)

Contextual Consideration for Implementation

In this last category of 12 questions, the NCDDR points out that outcomes of dissemination efforts are influenced by the scope of the dissemination plan, as well as by follow-up activities. For example:

- Does your project design clearly describe dissemination goals, strategies, and expected outcomes? (Scale 0: Don't Know; 1: No; 2-4 [blank], 5: Yes)
- Does your project periodically sample to determine the manner in which users learn about the availability of your project's information? (Scale 0: Don't Know; 1: No; 2-4 [blank], 5: Yes) (Scale 0: Don't Know; 1: No; 2-4 [blank], 5: Yes)
- Does your project provide a variety of ways for potential users to discuss implementation of project information with project researchers of content experts? (Scale: 0: Don't Know; 1: No; 2 [blank], 3: Somewhat; 4 [blank]; 5: Yes)

Discussion

How well are OERI-funded initiatives evaluating the impact of dissemination efforts? An initial scan suggests not much beyond the level of "bean counting", at this point. To be fair, however, bean counting is a necessary first step towards establishing the impact of dissemination. But more can be done. In order for more substantial evaluations to take place, project evaluators, project directors, and OERI personnel can work together in several ways to restructure dissemination and evaluation practices. First, a dissemination plan and strategy detailing what activities will take place when can be submitted as a Year 1 deliverable. In this way, all project administrators can focus early on dissemination's central role. Moreover, mechanisms to collect critical data can be put into place a priori, which would alleviate pressure on the evaluator at a project's conclusion. Second, the grant program itself can contain an expectation that some degree of information-sharing take place two years prior to a project's conclusion (e.g., Year 3 of a five-year program). In this manner, dissemination efforts can be piloted, evaluated and refined during the course of the project and be stronger and more effective in the later years. This would also allow evaluators the time, funds, and ability to focus on the outcomes of those dissemination efforts, rather than simply documenting the extent of dissemination activity. Third, absent the ability to consider dissemination impact outright or restructure the grant program, evaluators can make use of the NCDDR's criteria to determine how well and how thoroughly the project's dissemination framework has been articulated. The application of criteria like those of the NCDDR's to the OERI-sponsored project can then be included as part of the annual evaluation report.

Knowledge sharing is a traditional and vital component of the research cycle, the aspect of research that paves the way towards improved interventions. As evaluators, we are poised to assume the important role of improving the efficiency of educational technology initiatives' information sharing. We cannot do this on our own, however. In order to examine the process and outcomes associated with dissemination, project directors and administrators must also acknowledge and respond to the importance of tracking dissemination efforts. At the end of the day, evaluators, project personnel, target populations, and funding entities all benefit from examining dissemination efforts throughout a project's lifespan and not just by means of retrospective examinations that can be too late to exert any real impact.

References

Dodge, B. (1997). *Some thoughts about WebQuest.* Available online: http://edweb.sdsu.edu/courses/edtec596/about_webquests.html

Glik, D., Berkanovic, E., Macpherson, F., Strone, K. Ratner, D. & Connell Jones, M. (2000). Dissemination and utilization of an immunization curriculum for middle schools in California. *International Electronic Journal of Health Education*, *3*(2), 75-83.

National Center for the Dissemination of Disability Research [NCDDR] (1997). *Dissemination self-inventory*. Austin, Texas: Southwest Educational Development Laboratory. Available online: http://www.ncddr.org

Rogers, E.M. (1995) *Diffusion of innovations* 4th edition. New York: The Free Press.

Ryan, B. & Gross, N.C. (1943). The diffusion of hybrid seed corn in two Iowa communities. *Rural Sociology, 8,* 15-24.

San Diego City Schools (2000). *Triton dissemination efforts.* Retrieved June 1, 2001 from the World Wide Web: http://projects.edtech.sandi.net/dissemination/t_dissemination/index.htm

Saywell, D. & Cotton, A. (1999). *Spreading the word: Practical guidelines for research dissemination strategies*. Leicestershire, UK: Loughborough University.

Shrinkfield, A. J. (1983). Designing evaluations of educational and social progress by Lee J. Cronbach: A synopsis. In G.F. Madaus, M.S. Scriven., & D.L. Stufflebeam (Eds.), *Evaluation models: Viewpoints on educational and human services evaluation* (pp. 357-379). Boston, MA: Kluwer-Nijhoff Publishing.

8. Evaluator Contact Information

Below is the contact information for evaluators of projects that were cited in the Sourcebook. In parentheses after the project name is the funding source. TICG indicates Technology Innovation Challenge Grant and Star indicates Star Schools Project. More information on TICG projects can be found at www.ed.gov/Technology/challenge/. Information on Star Schools projects can be found at www.dlrn.org/.

ACT Now! (Anderson, IN) — (TICG)
Jerrell Cassady, Ball State University
TC250, Muncie, IN 47306
(765) 285-8500 / jccassady@bsuvc.bsu.edu

ACT Now! (Sweetwater) — (TICG)
Marcie Bober, Eleanor Lynch, San Diego State University
College of Education, 5500 Campanille Drive, San Diego, CA 92182-1182
(619) 594-0587 / bober@mail.sdsu.edu

Alliance+: Professional Development Through Technology — (TICG)
Mario Yepes-Baraya, Harcourt Educational Measurement
P.O.B. 6708, Lawrenceville, NJ 08648
(609) 734-5357 / m.yepes@att.org

Connections Project — (TICG)
Neal Topp, University of Nebraska at Omaha
Office of Internet Studies-6001 Dodge Street, Omaha, NE 68182
(402) 554-2435 / topp@UNOmaha.edu

Delaware Challenge Grant-Lightspan — (TICG)
Susan Giancola, University of Delaware
Newark, Delaware 19716
(302) 831-4437 / giancola@udel.edu

Education Future NOW — (TICG)
Tanya Suarez, Suarez and Associates
1008 Clarendon St., Fayetteville, NC 28305
(910) 678-8773 / tsuarez@fayettevillenc.com

Education Technology Improvement Plan — (TICG)
Dean Arrasmith, Northwest Regional Educational Laboratory
101 SW Main, Suite 500, Portland, OR 97204-3297
(503) 275-9557 / arrasmid@nwrel.org

Eiffel Project — (TICG)
Veronika Denes, Columbia Univ, Teachers College
NCREST Box 110, 525 West 120th Street, New York, NY 10027
(212) 678-4098 / denes@exchange.tc.columbia.edu

El Paso Collaborative for Academic Excellence — (TICG)
Myrna Gantner, University of Texas at El Paso
College of Education, Room 414, El Paso, TX 79968
(915) 747-5783 / mgantner@utep.edu

Generation WHY
Michael Coe, Northwest Regional Education Laboratory
101 S.W. Main St., Suite 500, Portland, OR 97204
(503) 275-9497 / come@nwrel.org

Iowa Distance Education Alliance — (Star)
Nancy Maushak, Texas Tech University
College of Education, P. O. Box 41071, Lubbock, TX 79409-1071
(806) 742-1997 / nmaushak@iastate.edu

Key Instructional Design Strategies — (TICG)
Maroba Zoeller, Allen Independent School District
601 E. Main St., Allen, TX 75002
(972) 727-0493 / maroba_zoeller@allenisd.org

LiteracyLink — (Star)
Jerome Johnston, University of Michigan
Institute for Social Research, P. O. Box 1248, Ann Arbor, MI 48106-1248
(734) 763-3079 / jerej@umich.edu

Louisiana Challenge Project — (TICG)
Sue Jackson, University of Louisiana
P.O.B. 42051, Lafayette, LA 70504
(337) 482-1119 / sujax@louisiana.edu

Maryland Electronic Learning Community — (TICG)
Winnie Mucherah, Baltimore Learning Community
6101 Breezewood Court, Greenbelt, MD 20770
(301) 345-0303 / wm62@umail.umd.edu

Mississippi Educational Television Project — (Star)
Stephen Merker, Merker Associates
2517 Top Hill Road, Louisville, KY 40206
(502) 895-4545 / (502) 895-4544 / merker@louisville.edu

San Antonio Technology in Education — (TICG)
Richard Diem, University of Texas at San Antonio
501 W. Durango Blvd., San Antonio, TX 78207-4415
(210) 458-2765 / rdiem@utsa.edu

Teacher Led Technology Project — (TICG)
Linda Toms Barker, Berkeley Policy Associates
440 Grand Ave, Suite 500, Oakland CA 94610
(510) 465-7884 / linda@bpacal.com

Trails Project, The — (TICG)
Judy Pfannenstiel, Research and Training Associates
9209 West 110th Street, Overland Park, KS 66210
(913) 4518117 x1236 / jpfannenstiel@rtainc.com

Triton Project — (TICG)
Marcie Bober, San Diego State University
College of Education, 5500 Campanille Drive, San Diego, CA 92182-1182
(619) 594-0587 / bober@mail.sdsu.edu

TV411
Jerome Johnston, University of Michigan
Institute for Social Research, P. O. Box 1248, Ann Arbor, MI 48106-1248
(734) 763-3079 / jerej@umich.edu

Virtual High School — (TICG)
Andrew Zucker, SRI International
1611 North Kent St., Arlington, VA 22209
(703) 247-8523 / zucker@wdc.sri.com

9. Measurement Appendix

Reproductions of instruments used in a number of Star Schools and Challenge Grant evaluations are contained in a collection of PDF files available on the Web site of the Distance Learning Resource Network (www.dlrn.org). A directory appears below

Teacher Surveys

Contents	Project	PDF File Name	PDF Page
1. Teacher Survey of Technology Use/ Integration	PUHSD/Global Connections	SbkAppx1 (1.5 Mbyte)	1-3
2. Teacher Survey of Technology Use/ Integration	ACT Now! Sweetwater Union H.S.		4-10
3. Teacher Baseline Questionnaire	Teacher Led Technology Challenge		11-26
4. Teacher Follow-up Questionnaire	Teacher Led Technology Challenge		27-36
5. Teacher Online Survey of Technology Use	Utah Technology Awareness Project	SbkAppx2 (2.5 Mbyte)	1-6
6. Teacher Self-Rating of Technology Skills and Use of Technology in the Curriculum	Milken Exchange		7
7. Teacher Survey of Skills, Classroom Practices, and Technology Beliefs	ACT Now! Sweetwater Union High School		8-15
8. Survey of Instructional Practices and Beliefs	Four Directions Challenge Grant		16-27
9. Status and Impact of Technology Showcase on Educational Technology: Current and Past Winners	Iowa Star Schools		28-33

Classroom Observation Tools

1. Classroom Observation Instrument for Elementary/Middle Schools	Teacher Led Technology Challenge	SbkAppx3 (3.2 Mbyte)	1-11
2. Classroom Observation Summary	Baltimore learning Community		12-15
3. Classroom Observation Instrument	Connections Project		16-22

Site Visit / Case Study Tools

1. Case Study Instruments: Classroom Organization, Authenticity, Teacher Practices, Student and Teacher Interviews	ACT Now! Sweetwater Union H.S.	SbkAppx4 (1.2 Mbyte)	1-18
2. Principal/Director Questionnaire	Eiffel Project	SbkAppx5 (2.0 Mbyte)	1-10
3. Technology Integrated Unit of Study Rubric for Teachers	Education Future NOW		11
4. Technology Coordinator Questionnaire: Estimate Project Effects	Eiffel Project		12-22

Student Technology Skills Surveys

1. Technology Skills Checklist, Grades 3-5	Teacher Led Technology Challenge	SbkAppx6 (2.2 Mbyte)	1-4
2. Technology Skills Checklist, Grades 4-6	Teacher Led Technology Challenge		5-9

Surveys for Pre-Service Teachers

1. Survey: Information Technology In Teacher Education	Milken Exchange on Education Technology/ ISTE	SbkAppx6 (continued)	10-15
2. Teacher Technology Survey (Pre-Service Teachers)	University of Northern Iowa		16-22

Adult Literacy Measures

1. Various Instruments Used in the Assessment of Workplace Essential Skills	LiteracyLink	SbkAppx7 (1.1 Mbyte)	1-49
2. Various Instruments Used in the Assessment of TV411	LiteracyLink		50-139