CHAPTER 3

Large-Scale Quantitative Research on New Technology in Teaching and Learning

RONALD E. ANDERSON
UNIVERSITY OF MINNESOTA, USA

Introduction

Large-scale empirical studies, especially in cross-national contexts, offer great payoff for the understanding of new literacies and new media, especially those related to teaching and learning. Without scientific sample and standardized measurement, generalization to larger populations is impossible or weak at best. Statistical designs with random-sampling techniques make it possible to generalize to huge populations of learners with relatively small samples and known estimates of error. Standardization of assessment instruments and coordination of procedures for collecting data make it possible to compare national and regional education systems, as well as demographic groups. Longitudinal and panel designs make trend analyses possible, because measures are repeated and statistical variations can be minimized. Experimental designs make it possible to isolate causal processes, because predictive conditions can be controlled and isolated to determine if they affect one or more student outcomes.

These methods, however, tend to be demanding in terms of cost, time, and collaboration. In addition, these types of research on new media issues face unique problems such as rapidly changing technologies. This chapter will review theoretical frameworks; describe the national and international large-scale studies of new media that have been conducted; review some major findings from these studies; discuss methodological issues and options; and discuss challenges and opportunities for such research in the future.
The scope of the studies discussed in this chapter encompasses both student assessments and surveys of utilization of new media, including pedagogical practices of teachers using information and communication technology (ICT). With regard to assessment, we include both studies that assess ICT-related new literacies and selected studies that provide for investigation of the impact of ICT utilization on learning in other subjects, such as mathematics or science.

One evident change during the past three decades has been the terminology for the technology that delivers the new media. Whereas during the 1980s, the principal new media was called “computers,” “information technology,” or “IT,” by the late 1990s, educators in most countries began to refer to it as “ICT” for “information and communication technology.” The integration of the Internet and multimedia with computer technology made many feel more comfortable with calling it “ICT.” In some countries, however, most notably in the United States, educators refer to it as “information technology,” “IT,” or simply as “technology.” In this chapter, the acronyms ICT and the word technology will be used synonymously.

This chapter is divided into four main parts or sections. The first section discusses theoretical issues, highlighting conceptualizations in relevant, large-scale studies. The second part describes some key features of each relevant study, ending with one or two illustrative findings from the study. The third section points out common methodological solutions in these types of studies. In the final section, speculations are made about the future and the research that will be needed.

Conceptual and Theoretical Approaches

Many of the chapters of this volume ask the theoretical question, “What are the new literacies?” In this chapter the theoretical question asked is “What should be the priorities when planning large-scale studies of new literacy?”

Relevant conceptual and theoretical work focuses on two areas: one is a measurement or assessment and the other is contextual and causal frameworks. The former is generally concerned with criterion or predicted variables, typically test scores but sometimes attitudinal and behavioral factors. The latter focuses on the interrelation of the predictor variables and their relation with the criterion or predicted variables. Within the discussion of each of these two conceptual tasks the approaches taken in relevant, major studies will be outlined. First, however, a theoretical approach will be outlined for addressing the assessment requirements of a broad spectrum of new literacies.

Toward a Theory for Assessing New Literacies

A project to systematically assess new literacies, particularly large-scale studies, must narrow, or delimit, the scope of the assessment in various ways. If one is interested in a very new and novel media, the choices are likely to be very limited. However, if one is interested in a broad scope of ICTs, then it is
necessary to prioritize components and dimensions of the full range of potential content, knowledge, and skills that could be assessed.

In order to delimit the new literacies assessment domain the following dimensions of digital tools, and their associated skills, are proposed:

1. Difficulty/Level: the amount of time and requisite skills required to learn the subject matter or performance. It is assumed that grade level and age are strongly associated with difficulty.

2. Does the content require hands-on or performance assessment or are other forms of assessment feasible?

3. Complexity of problem solving required.

4. The extent to which the information and/or knowledge ranges from highly concrete, explicit information to subjective or tacit knowledge.

5. The length of time that the media or technology has persisted. Has it been in place for a long time, avoiding obsolescence or is it new and novel?

6. Does the task for the student outcomes being assessed require a particular media or ICT or can it be completed manually in a reasonably comparable amount of time without the media or ICT?

Dimensions 2, 5, and 6 are defined in terms of two states or conditions; however, one should assume that these states define polar opposites and a wide range of intermediate states fall in between.

Another assessment condition will be treated as a design factor rather than as a planning dimension and that is the role for which the new literacy is considered relevant or critical. The principal roles that underlie new literacy tasks are productivity or employment roles; learning or student roles; and civic or citizenship roles. Of course, there are other roles such as social roles and leadership roles for which new literacies are relevant, but these roles are generally not salient enough for large-scale assessments. The researcher will initially select one or more roles and then determine what knowledge and skills pertaining to new literacy are relevant to that role or roles.

Given resource constraints, it is not feasible for any one research project to include assessment items for each point along each of these six dimensions or their underlying continua. With respect to difficulty, the researcher will typically select one to three age groups and then target tasks toward the large majority of those within the age group(s). With respect to the other dimensions, decisions are too often made implicitly or without sufficient planning. There are good reasons for choosing a specific decision point on some of the dimensions. However, before discussing this possibility, the major projects defining and measuring computer or ICT literacy will be described first.

Assessment Framework Projects

Comparative assessments typically begin with an assessment framework, which is intended to define the content universe from which test items will be
sampled. It also identifies the specific facets or domains and subdomains of the content to be assessed. Key elements of the assessment frameworks will be summarized for each of the large-scale projects that attempted to define aspects of computer or ICT literacies. Many, but not all, of these projects actually attempted to measure technology-related literacies in the field. Empirical details of those studies will be described in the next section of this chapter.

The Minnesota Computer Literacy Assessment (MCLA). In preparation for the 1979 MCLA, the investigators developed the first conceptual framework for the measurement of skills, knowledge, and attitudes relevant to computer utilization by students (Johnson, Anderson, Hansen, & Klassen, 1980). Both the knowledge domain of this framework and the paper-and-pencil test were divided into three subdomains: (a) knowing basic computer concepts; (b) knowing applications and their impact; and (c) understanding and reading simple algorithms (Anderson & Klassen, 1981). While the definition of computer literacy was hotly debated, the attitude and knowledge measures from this study provided the foundation for most of the hundreds of studies on computer attitudes or computer literacy during the decade that followed (Anderson, Klassen, & Johnson, 1981).

The ETS computer competence study. In the early 1980s, the most hotly contested technology issue in education was whether or not computer literacy should include and emphasize computer programming. The programming proponents won out, and in 1986 the Educational Testing Service (ETS) conducted a national assessment of computer programming, calling it “computer competence” (Martinez & Mead, 1988). The study was done under the auspices of NAEP (National Assessment of Educational Progress). While their domains were defined in terms of (a) knowledge and attitudes, (b) computer applications, and (c) computer programming, the principle emphasis was on programming.

The IEA computers in education (CompEd) study. The IEA CompEd study was the first international, technology-related large-scale survey and assessment. Nearly 20 different countries were involved in one or more segments of the student survey between 1988 and 1993 (Pelgrum & Plomp, 1991). For the main student assessment in 1992, the project developed an instrument called the FITT (Functional Information Technology Test). The conceptual domains defined for this test were three-fold: (a) knowing basic computer concepts; (b) knowing applications and their impact; and (c) understanding and reading simple algorithms. An optional, short, and separate computer programming test was developed as well. It is noteworthy that even though this assessment framework was developed more than 10 years after the framework of the MCLA, the categories were quite similar. Micro and personal computers had
emerged during that time, but most of the categories and priorities of information technology and information science remained the same.

The NETS (National Educational Technology Standards) project. Because of the demands from schools and teachers, the NETS (National Educational Technology Standards) project was organized by ISTE (International Society for Technology in Education) in the mid-1990s. The resulting standards were published by ISTE (1998), and they remain widely used by teachers, schools, districts, states, and professional associations. The student NETS are embedded within ISTE’s NETS-T (standards for teachers) and NETS-A (standards for school administrators). In principle, all teachers and administrators are supposed to meet the student standards as well. NETS is used by NCATE (National Council for Accreditation of Teacher Education) for alignment of standards for technology facilitators and technology leaders. Many assessments have used NETS as guides. For example, North Carolina requires that all 8th-grade students pass a test that is somewhat based on NETS.

The scope of the standards is suggested by the six topics in Table 3.1. Associated with each standard are several criteria for how the standard should be used. The rightmost column gives an illustrative criteria for each standard.

In addition to the standards in Table 3.1, the NETS standards manual includes (a) performance indicators, e.g., “choose an appropriate tool and tech resources for a given task”; and (b) curriculum scenarios with learning activities, e.g., in social studies unit for 9 to 12, students are to create a Web site on population growth after conducting various types of research; and (c) assessment guidelines, e.g., sample rubrics might be provided for a given criterion or standard.

NETS was not designed to be a measurement framework for a research project or assessment, so by themselves the NETS does not constitute a complete conceptual framework. It is not even always explicit about what students

<table>
<thead>
<tr>
<th>No.</th>
<th>Standard topics</th>
<th>Sample criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Basic operations and concepts</td>
<td>E.g., Students are proficient in the use of technology.</td>
</tr>
<tr>
<td>2</td>
<td>Social, ethical, and human issues</td>
<td>Students practice responsible use.</td>
</tr>
<tr>
<td>3</td>
<td>Technology productivity tools</td>
<td>Students use technology tools to enhance learning.</td>
</tr>
<tr>
<td>4</td>
<td>Technology communications tools</td>
<td>Students use a variety of media and formats to communicate effectively to multiple audiences.</td>
</tr>
<tr>
<td>5</td>
<td>Technology research tools</td>
<td>Students use technology to locate, evaluate, and collect information from a variety of sources.</td>
</tr>
<tr>
<td>6</td>
<td>Technology problem-solving and decision-making tools</td>
<td>Students employ technology in the development of strategies for solving problems in the real world.</td>
</tr>
</tbody>
</table>

Adapted from ISTE (1998).
should know. Nonetheless, it is a useful guide for planning instruction and curriculum. Most importantly, it is used very widely by teachers, and it is updated and expanded every year.

The IT fluency project. The IT Fluency project was sponsored and administered by the National Research Council (NRC; 1999) of the United States, and the report was published by the National Academy Press. A panel of mostly computer scientists was convened as the starting point for the conceptualization. Their framework consisted of a number of categories of IT fluencies within each of three major domains: (a) IT concepts; (b) IT skills; and (c) intellectual capabilities. The first two domains were quite similar to the concepts and applications dimensions of earlier studies. However, the “intellectual capabilities” domain contained some rather complex and challenging topics expressed as behavioral objectives, specifically, “manage complexity,” and “think about IT abstractly.” Consequently, the implicit definition of IT fluency was quite advanced, and it would be unlikely that anyone would be judged to be IT fluent by these criteria unless he or she had taken several college-level computer-science courses at a minimum.

The IEA SITES student-assessment project. The IEA SITES project developed a knowledge-management framework for assessing ICT-related skills (Anderson & Plomp, 2002). Based on a synthesis of numerous statements about desired 21st-century student outcomes, a list of knowledge-management competencies was constructed and these appear as the rows of the framework in Table 3.2. The columns in Table 3.2 represent various domains or categories of ICT tools. This typology of ICT tools is not based on IT industry classifications but on what are considered the most useful ICT applications for teaching and learning. The typology borrows heavily from Jonassen’s (2000) classification of Mindtools. The cell entries are desired student outcomes that could be used as evidence for the associated knowledge-management competency. The desired outcomes in the cells presume that the student is using one or more ICT tools in the associated row. It should be evident that this framework is intended for a performance assessment where the student has specific software applications available. In some of the cells, a specific software application, e.g., SIMCALC, is given to illustrate the type of tool available.

Some of the project problem-solving tasks in this framework are presumed to be a shared or collaborative task. Other knowledge-management competencies could be defined as collaborative, but collaborative performance tasks are even more challenging to construct than performance tasks for individuals. The assessment based on this framework was approved by the IEA, but sufficient funding was not found to carry it out. It is hoped that some performance assessments utilizing this type of conceptual framework will be implemented in the future.
<table>
<thead>
<tr>
<th>ICT Knowledge Management Competency</th>
<th>Knowledge construction tool kit and data environment</th>
<th>Semantic organization tools</th>
<th>Dynamic modeling tools</th>
<th>Interpretation tools, e.g., visualization and search tools</th>
<th>Communication and presentation tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge construction</td>
<td>Project products: reports &amp; presentations</td>
<td>Arguments and conclusions</td>
<td>Model design with conclusions</td>
<td>Project products: reports and presentations</td>
<td></td>
</tr>
<tr>
<td>Critical thinking: analyze, interpret data, evaluate evidence</td>
<td>Formulate critiques and conclusions</td>
<td>using online concept maps, represent evidence</td>
<td>Scenario simulation (see e.g., in Bennett, 1999)</td>
<td>Using data mining tools to drill down to highly granular information</td>
<td></td>
</tr>
<tr>
<td>Collaborative projects and complex problem solving</td>
<td>Derive solutions for problems</td>
<td>Derive solutions for problems</td>
<td>Use an optimization model for decisions</td>
<td>Interpret data using visualization tools</td>
<td></td>
</tr>
<tr>
<td>Effective presentations and discourse</td>
<td></td>
<td></td>
<td></td>
<td>Develop persuasive Web-based presentation</td>
<td></td>
</tr>
<tr>
<td>Find, assemble, restructure knowledge</td>
<td></td>
<td>Make correct inferences using, e.g., SIMCALC</td>
<td>Web search, compile, and analyze</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Understand principles including secondary effects</td>
<td>Discuss social implications of using above technology</td>
<td>Discuss social implications of using above technology</td>
<td>Discuss social implications of using above technology</td>
<td>Discuss social implications of using above technology</td>
<td></td>
</tr>
</tbody>
</table>

Note 1. The table structure was taken from Anderson and Plomp (2001). The table contains some changes to improve clarity.
The ETS ICT literacy project. ICT Literacy standards were defined by the ICT Literacy Project at the Educational Testing Service (International ICT Panel, 2002). The core construct in its framework of standards was called ICT Proficiency. This was defined in terms of five information management skills: (a) access, (b) manage, (c) integrate, (d) evaluate, and (e) create information. While these five types of information management skills are at the core of their conceptualization, they also claim that ICT Literacy necessarily encompasses Cognitive Proficiency and Technical Proficiency as well. The cognitive domain consists of foundational skills of general literacy and problem solving. The technical domain parallels what is often called technological literacy, and subsumes considerable knowledge of hardware, software, and networking.

The Educational Testing Service had hoped for funding to use its framework in an international assessment. Even though the funding was not forthcoming, their framework has been used in the design of instructional modules.

The PISA ICT literacy project. The ETS ICT Literacy framework was adopted by PISA planning groups; however they added “communication” to the five knowledge-management skills. Presently, PISA has not determined if this will be the final conceptual framework to be used in its 2009 assessment.

Summary of assessment frameworks. Developing conceptual models for what should be taught and tested related to the new media has been very challenging. Perhaps the main reason is that the new media are still rather novel and their capabilities change quite rapidly. Yet, the assessment frameworks changed very little for 20 years, up until 2000 and later. Probably this stability occurred because ICT functions remained fairly constant.

In the nearly 30 years since new media literacy studies began, editing and composing text arguably remains the highest priority ICT-based task for the educated citizenry. At the same time, electronic networks and e-mail have been emerging and have yielded new functions that students and others should become familiar with.

During the last 3 decades, the science of both text and graphical processing has made major strides. These strides combined with the Internet have greatly increased the salience of information and knowledge management. Recent assessment frameworks for new media reflect this salience.

The history of the assessment frameworks reviewed in this section so far hint that future framework development must continue to balance change with continuity. It is a delicate balance, which perhaps can be resolved by concentrating on the goals of the endeavor.

Contextual and Causal Frameworks
The other conceptual issue most challenging for large-scale assessment-related studies is a contextual or structural one: what are the relevant contextual
factors for understanding or predicting teaching and learning processes and how are these contextual factors best structured? If the project is conceived as a qualitative one, the challenge is likely to be defined in contextual terms; whereas if it is quantitative, then the framework is likely to be defined structurally, especially in terms of how the contextual factors are interrelated in predicting performance on the main outcomes of interest.

**Contextual frameworks.** These conceptual questions generally cannot typically be answered succinctly. The substantial complexity of the categories and variables involved are often summarized graphically or given a visual representation. Figure 3.1 is a concentric circle diagram typically used to show expanding contexts of influence. Cole (1996) provided the inspiration for this framework, which was used in the early planning phases of the qualitatively oriented SITES case-studies project. In this type of model, each larger circle is assumed to influence the circles within it. The slices represent factors like the curriculum that impinge on or across all contextual levels. The framework (or model) is not intended to be precise in that any given outer circle might have direct influence on more than one concentric circle within it. The labels in the

![Figure 3.1 Sample conceptual framework: Contextual aspects in the SITES case-studies project.](image-url)
top half of the center column define the concentric circles and the labels in the bottom half suggest the types of data-collection methods that are likely to be most appropriate for that level of the circles.

This type of framework especially assists qualitative study design and analysis, because it helps to point out a huge number of characteristics that may play a role in understanding the central phenomenon or outcome at the very center of the diagram, which in this case is student learning. It also implicitly points out how many contextual layers need to be considered. This model (Figure 3.1) was especially helpful at the initial planning phases of the IEA SITES qualitative case studies project called Module 2. It helped to reinforce our initial presumptions that we knew little about the influence of a multitude of factors on learning, and we needed to design the study so that the researchers would be open to discover any set of interrelationships.

Causal frameworks. While the model just described is a contextual one, a structural or causal model may be a much more appropriate framework, especially if the project is quantitatively oriented. Figure 3.2 is a hypothetical causal model that might well be applied to the framework of an assessment of knowledge and skills regarding new literacy. The model helps to identify how various variables or variable blocks are (or are not) related to a student new-literacy measure.

A causal model helps outline the implicit or explicit links among the key elements or forces. The boxes represent concepts and their associate variables; lines represent relationships; a line with an arrow stands for a causal effect; and a line with two arrows indicates mutual causation, which is also called a reciprocal relationship. A causal model or framework will not be useful if all concepts or factors impinge on all others. In such an event, every box in the diagram would have a two-way arrow directed at every other box, but the effect of any one variable on another could not be estimated precisely using mathematical procedures. In statistical terms, it would be called a nonspecified model.

The model depicted in Figure 3.2 is a more useful model in that some refining conceptual assumptions have been made. These include assumptions such as the following: that family/community affect classroom processes and student new literacy indirectly only, by first having an effect on student attitudes or school characteristics; that school characteristics and family/community are not affected by either student or classroom variables; and so forth. These assumptions may be wrong, but they do provide hypotheses that can be tested and found to be either confirmed or not by an appropriate set of data. If the measures or indicators well represent the concepts given in the boxes, then this type of modeling and analysis can yield important knowledge about specific factors that predict student performance and how these factors are interrelated.
To date, the large-scale studies attempting to predict new literacy (computer literacy or ICT literacy) have been disappointing. This may be due to the poor quality of the measures of new literacy. More likely, it is a consequence of the descriptive emphasis given in the data analysis of these projects. For whatever reason, the researchers tended to stop short of extensive causal analysis in these studies.

**Blended frameworks.** The closest many large projects come to causal modeling is in constructing what one might call a “blended framework.” The blended model or framework combines the contextual and causal approaches, diagramming relationships of both types. A sample blended model is given in Figure 3.3, which was the framework derived by the IEA SITES case-studies project.

As you can see by this diagram, it has a very large number of concepts and relationships incorporated into the framework. This achieves the objective of making a list of all the relevant variables and showing how subsets of these variables might be interrelated. While the blended framework has the appearance of technical sophistication, it gains relatively little over the more primitive contextual framework depicted in Figure 3.1.

Because of many of the labels and relationships that are not at all self-evident, an explanation of such a framework may require several long chapters. Even then, the categories and relationships may not be precisely defined. Nonetheless, such frameworks have an important added value: once the researchers have exerted the time-consuming and tedious effort required to construct such a model, they are likely to be ready to proceed to undertake
a qualitative study such as a case-studies project. Furthermore, having given some thought to the causal linkages by constructing a blended framework, it may be easier to produce a variety of causal models each dealing with a more narrow scope of the full domain.

**Relevant Large-Scale Empirical Studies**

*The Minnesota Computer Literacy Assessment (MCLA)*

Nearly 30 years ago, the first large-scale study of new literacies was conducted in Minnesota. The Minnesota Computer Literacy Assessment, conducted by Anderson, Klassen, Krohn, and Smith-Cunnien (1982) at the Minnesota Educational Computing Consortium (MECC), in which 3,600 randomly selected 8th- and 11th-grade students in Minnesota schools were surveyed and assessed. Anderson, Klassen, and Hansen, (1981) also surveyed 1,000 randomly selected mathematics and science teachers for the 8th and 11th grades and identified some of the factors that facilitated and curtailed the adoption of computers in their classrooms.

At the end of the 1970s, punch cards were still a popular medium for computer utilization. One might be tempted to exclude these early stud-
ies from new media (or new literacies) research because of the primitive technology in use then; however, at that time timesharing terminals and microcomputers were already being used extensively to deliver computer assisted (CAI) and computer managed instruction (CMI). During the 1960s and early 1970s, CAI and CMI were largely limited to what was called “drill-and-practice” instruction, but many tutorials, some of which used graphical displays, were utilized by the end of the 1970s in secondary and postsecondary schools. By that time, databases and word processors also were commonly used in education and their use was taught in a number of classrooms. Microcomputers had even been acquired for a few schools and some households.

To provide a glimpse of the MCLA test, here is one of the representative test items in the MCLA:

A Newspaper publisher has the following information about subscribers stored in the computer. They are name, address, and renewal date. How would you arrange the information to be most useful to the delivery person?

a. ordered listing by address
b. ordered listing by renewal dates
c. alphabetical listing of streets
d. ordered listing by zip code
e. I don’t know

Only 30% of eighth graders and 35% of eleventh graders answered this test item correctly. The poor performance on this item reflects the fact that most students at that time had not been exposed to either the use of a database program or the principles of algorithmic logic.

What is noteworthy about this item, and others like it, is that it is a measure of information literacy, specifically the ability to organize information in such a way as to make it readily accessible and analyzable. Most of the literature on information literacy and new literacies did not begin to appear until the late 1980s and early 1990s (cf., Spitzer, Eisenberg, & Lowe, 1998; Hawisher & Selfe, 2000). New literacies may be deictic, but some elements have been highly persistent. Text editing and organizing information by sorting are technologies that have been critically important for the past 30 years. Yet, because they are “old” technologies, researchers may neglect them.

The test item is additionally interesting because girls were 5% more likely than boys to answer it correctly. The same pattern occurred across some other items associated directly or indirectly to programming logic. This is indicative of how women sometimes have an advantage over men in tests of information literacy. This is especially true with test items requiring procedural logic and skill and/or familiarity with general productivity tools (Anderson, 1987).
The ETS Computer Competence Study

The 1986 study of “computer competence” by the Educational Testing Service (ETS) was mentioned earlier, and it was noted that the assessment domain was divided into three sections: (a) knowledge and attitudes, (b) computer applications, and (c) computer programming. The principle emphasis was on programming.

The study was a massive undertaking. A total of 632 test items were administered in a block design to well over 25,000 students in grades 3, 7, and 11 (Martinez & Mead, 1988). From this study, we know that 90% of the students in those grades had used computers and about 50% were getting some kind of instruction about computers in school. Over a third of the students had a computer at home, even though the home computer was then only a few years old. One of the principal findings of the study was that the exposure to computers at home was equally as significant in predicting students’ computer competence as what they were learning in school. A major digital gap was evident in that they found a major advantage in computer competence for students who had a high SES and parents with a college education, who were not racial minorities.

By the late 1980s, packaged productivity software such as word processors, spreadsheet programs, and database programs had become available for all the major microcomputers or personal computers, as they came to be called. This undercut the validity of arguments for universal instruction in computer programming. Meanwhile the term “computer literacy” had come to be used to refer to the skills needed for using computers effectively with productivity-software tools. By the time the results of the study were released, computer programming was no longer given much attention, which accounts for the study having had relatively little impact on education. Ironically, with the introduction of the Internet, the need for people to learn computer programming has arisen again. Some now argue that students should be introduced to, and tested in, HTML and XLM, which are similar to traditional programming languages.

The IEA Computers in Education (CompEd) Study

In 1995 the first large-scale, international study of technology in education was proposed by Tjeerd Plomp and associates at the University of Twente in the Netherlands and approved by the IEA (International Association for the Evaluation of Educational Achievement). The study had two separate stages—one collecting data in 1989, and the second collecting data in 1992. Twenty-two countries participated in the first stage, and 12 countries completed the second stage.

During the 1989 (Stage 1) study, surveys were conducted in each of three types of schools: (a) elementary schools, (b) lower secondary, and (c) upper secondary (Janssen Reinen & Pelgrum, 1993). Within each school sample,
questionnaires were completed by the principal, computer coordinator, and several teachers. Sample surveys were completed in each of the 22 participating national systems: Austria, Belgium (Flemish), Belgium (French), Canada (BC), China, France, Germany (Federal Republic), Greece, Hungary, India, Israel, Italy, Japan, Luxembourg, Netherlands, New Zealand, Poland, Portugal, Slovenia, Spain, Switzerland, and the United States (Pelgrum & Plomp, 1993).

During the second stage in 1992, the surveys of school principals and computer coordinators were repeated, but focused on the added component of student assessment. The student questionnaire included attitude and background questions, as well as test items, and was administered to students in one randomly selected classroom within each sampled school. The countries participating in the 1992 study were Austria, Bulgaria, Germany, Greece, Italy, India, Israel, Japan, Latvia, Netherlands, Slovenia, Thailand, and the United States.

In 1992, the CompEd study collected data from over 75,000 students, 5,000 schools, and 10,000 teachers (Pelgrum, Janssen, Reinen, & Plomp, 1993). Each country surveyed a scientifically selected, representative sample of schools and a representative sample of students and selected teachers within schools. With the exception of India and South Africa, where only one region of the country was surveyed, and Greece, where only computer-using schools were sampled, the countries obtained samples representing all of the schools, both academic and vocational. The only excluded schools were those offering only special education or those not serving the targeted grades. The three targeted grades (student populations) were (a) the grade in which the modal age of students was 10 years, (b) the grade in which the modal age of students was 13 years, and (c) the next-to-the-final year of upper-secondary school. In most countries, this resulted in the selection of grades 5, 8, and 11.

This FITT test that was mentioned in the first part of the chapter was bundled with the student questionnaire and administered as a paper-and-pencil task. The study yielded many indicators of the quality of the role of technology across educational systems. At the school level, the indicators included hardware and software availability, location of equipment, and organization of computer support. At the teacher level were measures of teacher training, experience, and classroom practice. At the student level were indicators of knowledge, attitudes, exposure to different learning applications, and the use of home computers for educational purposes. Here are samples of the findings:

1. The number of computers in most countries rose rapidly during the 1980s and early 1990s, but by 1992 the number of computers in typical schools was only enough to allow one class at a time to use computers. Many of the computers were quite primitive. Over 50% of the computers in U.S. schools were Apple II computers (Anderson, 1993).
2. The most common use of computers in learning at that time was occurring in conjunction with courses on informatics or other courses for students to learn how to use computers (Pelgrum & Schipper, 1993). However, this was not the case in the United States where science, math, and language-arts teachers were expected to teach ICT skills as students needed them.

3. On the FITT test, which is best viewed as a test of practical computer knowledge, the Western European students in Austria, Germany, and the Netherlands had the highest scores. (See Table 3.3.) Eleventh-grade students in Latvia and Slovenia were also relatively high. On the other hand, students from Bulgaria and 8th-grade Japanese students had the lowest scores. In between these two extremes were students in the United States (Plomp, Anderson, & Kontogiannopoulou-Polydorides, 1996).

The relatively poor performance of Japanese and American students was a surprise to many due to the emphasis on technology in these countries. Japan was historically slow to put computers in their classrooms until a few years after this assessment. Schools and educators in the United States had been emphasizing the integration of computers into education for several years, but the teaching of computing skills had generally been neglected.

American educators tended to believe that students learned IT skills at home and that it was better for teachers of all subjects to be responsible for teaching IT skills only as needed. Most European schools, however, started introducing computer-related (informatics) courses, often on a required basis, in the late 1980s (Brummelhuis, A.C.A. t., 1993; Collis et al., 1996). Ironically, the emphasis on technology integration within American education has resulted in the relative absence of opportunities for elementary and secondary students to obtain instruction in ICT skills.

<table>
<thead>
<tr>
<th>Educational system</th>
<th>8th Grade</th>
<th>11th Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>69</td>
<td>86%</td>
</tr>
<tr>
<td>Germany</td>
<td>69</td>
<td>—</td>
</tr>
<tr>
<td>Netherlands</td>
<td>67</td>
<td>—</td>
</tr>
<tr>
<td>United States</td>
<td>61</td>
<td>65</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>51</td>
<td>62</td>
</tr>
<tr>
<td>Japan</td>
<td>49</td>
<td>65</td>
</tr>
<tr>
<td>Latvia</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Slovenia</td>
<td>70</td>
<td></td>
</tr>
</tbody>
</table>
Teaching, Learning, and Computing (TLC)

The Teaching, Learning, and Computing (TLC) study in 1998 surveyed principals, technology coordinators, and teachers from a national probability sample of all elementary and secondary schools in the United States and from two large targeted or purposive samples of schools: (a) high-end technology-using schools and (b) schools participating national and regional educational reform programs. Results were obtained for about 650 schools in the probability sample and 470 in the purposive samples. From almost all of these schools, questionnaires were received from the principal, the technology coordinator, and three to five teachers. Further details on the study methodology can be obtained from Becker and Anderson, 1998.

While the study was not designed to investigate new literacies, some questions were included regarding teachers’ skills with using technology. Their technological prowess was interrelated with beliefs about teaching and several other factors, such as the quality of technology support, were found to be closely correlated (Dexter, Anderson, & Ronnkvist, 2002).

Perhaps the most important findings of this study pertain to teachers and their participation in communities of practice (CoP). Participation in CoPs made it much more likely for teachers to successfully implement improvements in the quality of their instruction, especially if the school culture was in other ways supportive of their attempts to improve teaching (Becker & Riel, 1999).

ICT can play an important role in such instruction. The CoPs, sometimes called “professional communities,” operate effectively when teachers share ideas and feedback around their main goals in teaching. This is especially true when teachers’ goals are complex and not well understood. Two goals of this type examined in the TLC study were constructivist philosophy and belief in the importance of new literacies in teaching. Implementing effective innovations tend to be rare because of the many demands on teachers’ time and attention, but teachers with constructivist beliefs are more likely to use new media in their teaching and to innovate in implementing them.

The IEA SITES Study

The SITES (the Second International Technology in Education Study) study was initiated in 1997 by the IEA to investigate the role of ICT in education. The design of SITES followed the earlier CompEd model in some ways. Both studies had surveys at the school level of principals and of school-technology coordinators. The CompEd study had both a student assessment and a teacher survey and the proposed third module of SITES was to have a student assessment. The SITES study consists of three modules as summarized in Table 3.4.
The overall study design was approved by the IEA in 1997, and the survey data of module 1 were collected in 1998. The module 2 case study visits to the school sites were conducted during 2000 and 2001 and the report was released in 2003 (Kozma, 2003). Module 2006 scheduled their data collection for 2006 with the results to be released in late 2007. Each of the three modules will be described briefly in turn.

**SITES-module 1.** In 1998 data were collected using a questionnaire survey of principals and one of technology coordinators or their equivalents. Twenty-six countries participated by conducting these surveys in one or more of these three school levels: (a) primary, (b) lower secondary, and (c) upper secondary. As reported in Pelgrum and Anderson (1999), SITES-Module 1 produced findings on the extent to which ICT is used (and by whom) in education systems across the globe. These findings include,

- the extent to which education systems have adopted, implemented, and realized the results from objectives that are considered important for education in a knowledge society;
- teaching practices that principals consider to be innovative, important, effective, and satisfying;
- existing differences in ICT-related practices both within and between education systems and what lessons can be learned from this.

The findings on school Internet access were representative of the heterogeneous pattern of cross-national adoption of new ICT. Figure 3.4 shows that while 100% of the schools in Singapore and Iceland had access, some countries had only about a fourth of their schools connected. Most of the other countries had connected over 50% of their schools. What is so remarkable about this pattern is that even in countries that do not speak the dominant language of the Internet, English, most of their schools had been connected and many of the students were using the Internet in school. This rapid connection of schools to the Internet occurred within only about 5 years or less.

---

**Table 3.4 SITES Three Modules**

<table>
<thead>
<tr>
<th>Module</th>
<th>Time frame</th>
<th>No. of countries</th>
<th>Issue</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (SITES-M1)</td>
<td>1997–1999</td>
<td>26</td>
<td>What are the main trends?</td>
<td>Surveys of schools (principals and technology coordinators)</td>
</tr>
<tr>
<td>2 (SITES-M2)</td>
<td>1999–2003</td>
<td>28</td>
<td>What innovative teaching uses technology and what does it take?</td>
<td>In-depth case studies of innovative teaching in schools</td>
</tr>
<tr>
<td>3 (SITES-2006)</td>
<td>2002–2007</td>
<td>20 (est.)</td>
<td>What are teachers able to do with ICT to improve their learning?</td>
<td>Surveys of schools and teachers</td>
</tr>
</tbody>
</table>
On the other hand, Figure 3.4 shows the dismal nature of new media in education—a huge digital divide between the wealthier nations and the poorer ones. Except for Thailand and Bulgaria, developing countries were not included in this study. The Internet has largely passed over many of the countries in the world where electricity is a luxury. Even within wealthier countries like the United States where the digital divide has shrunk, there are still huge (and growing) disparities in the degree and quality of exposure to ICT both inside and outside of school.

**SITES-module 2 (case studies).** Nearly 30 countries conducted in-depth case studies during 2000 and 2001. The focus of this qualitative research is innovative pedagogical practices that use technology (IPPUT). Typical IPPUTs were classes that did large, group projects; participation in electronic field trips; art classes that switched to the use of painting software; and schools that used data-driven, teacher-based decision making. Each country was given the latitude of selecting its own criteria for case selection, but this had to be what was considered innovative in that country.

The main purposes of this qualitative study were to understand the essential characteristics of these practices, what sustains them, and what outcomes are produced. To accomplish this investigation, each case study describes and analyzes classroom-based processes and their contexts. These case studies were intended to provide policy analysts and teachers with examples of “model” classroom practices and offer policymakers findings regarding the contextual
factors that are critical to successful implementation and sustainability of these exemplary teaching practices using ICT.

The 28 countries participating in this module of SITES were Australia, Canada, Chile, China-Hong Kong, Chinese Taipei, Czech Republic, Denmark, England, Finland, France, Germany, Israel, Italy, Japan, Korea, Latvia, Lithuania, Netherlands, Norway, Philippines, Portugal, Russia, Singapore, Slovak Republic, South Africa, Spain, Thailand, and the United States. Each country conducted 4 to 12 case studies, and the total number of cases for analysis was 172. Some results and conclusions of this module can be found on the Web site: http://sitesm2.org. The final report of the project was published as a book by ISTE (Kozma, 2003).

One noteworthy finding was that the students used the Internet as part of nearly every innovative practice selected. Another finding of perhaps greater importance is that the students involved in these innovative pedagogical practices often engaged in activities that could be considered “knowledge management,” especially knowledge construction. Typically, such activities were called projects and included the tasks of searching, organizing, and evaluating knowledge. For instance, Germany’s first case study found that students “turned into providers of knowledge.” Portugal reported a case where the teachers wanted their students to be “constructors rather than receptors of mathematical knowledge.” In Norway, Australia, and the United States, to name a few such countries, cases were found where students worked collaboratively with ICT tools to complete large projects yielding diverse types of knowledge. Some of these projects involved shared activity and regular communication with students in other schools and countries.

Not all of the cases included collaboration, constructivism, or creativity in student activities. Imagine the traditional classroom that has received a donation of two laptops that sit in the back of the room to be used only by students who finish their assignments first. If you took a picture of the classroom from where the teacher stands, you would capture mostly children bending over their books at individual desks.

Now contrast that image with the students in the following photograph. These students were using MicroWorlds software to design and represent shape rotations in “Maths.” Here the teacher was acting as a facilitator or coach by asking questions. In this classroom, the students were working on their problems in collaborative groups.

The differences in these two types of classrooms were uncovered and elucidated in the IEA SITES study by using qualitative methods, especially classroom observation and interviews.

Survey indicators that captures such subtle contrasts can be constructed but not without extensive experience in the field settings and expertise in survey measurement. This study demonstrates the feasibility and value of using qualitative methods and case studies in an international context and on a fairly large scale.
SITES-2006. SITES-2006 was designed to build on these case study findings from the leading-edge classrooms of SITES-Module 2 (http://www.sites2006.net). Specifically, the school survey and the teacher survey questionnaires include indicators to identify the difference between innovative and typical learning contexts. The study was intended to determine the readiness of schools and teachers to provide a learning environment where students can participate in enriched pedagogical practices.

SITES-2006 also was originally designed to include an assessment of students, which was described in the previous section of this chapter. Sufficient funding, however, was not forthcoming for a student assessment, and it was dropped completely. The surveys of teachers and of schools (principals and technology leaders) went ahead as planned. Data collection proceeded in 2006 and the results are expected in 2007.

Countries participating in SITES-2006 sampled a minimum of 400 schools with the grade in which the majority of 14-year-old students were enrolled, which in most countries was the eighth grade. From each of the sampled and participating schools, attempts were made to obtain completed questionnaires from two mathematics teachers and two science teachers of students in the grade with the most students at age 14.

Despite highly diverse national educational systems around the world, almost every country has established policies regarding ICT in education (cf., Plomp et al., 1996; Plomp, Anderson, Law, & Quales, 2003). SITES in its various modules found many different approaches across countries in the way they take up the ICT challenge in education. Yet, there are common threads such as widespread and rapidly growing access to the Internet. It is particularly noteworthy that in so many countries the majority of educators believe in reform-like teaching practices, e.g., inquiry, project-based learning, and even
constructivist-oriented learning. Perhaps these norms of desirable pedagogical practice are largely a by-product of globalization and the exchange of new ideas via electronic networks.

**TIMSS.** Perhaps the best-known international assessment is the IEA’s Third International Mathematics and Science Study (TIMSS) conducted in 1995 and its later replications. It was partially repeated in 1999 as the TIMSS-Repeat study. And, in 2003 another wave of the study was conducted but renamed Trends in Mathematics and Science Study (TIMSS) with the intent that it would be repeated every 3 years. The 1995 study assessed students in grades 4, 8, and the final year of secondary school. In 1995, the TIMSS Video Study was conducted concurrently in Germany, Japan, and the United States. The original TIMSS study actually involved 45 countries.

The TIMSS study taught us that education in general, like the economy, is becoming global. Education, because it is the main source of human capital for a knowledge-oriented society, is increasingly critical to competitiveness in the international economy. The TIMSS experience shows that schools and their leaders across the globe want to learn from each other in order to improve their processes and outcomes. However, approaches used in one culture may not transfer to other systems without taking into account cultural differences and dependencies. For instance, the TIMSS Classroom Video Study found that Japanese teachers, compared to American teachers, were more likely to use methods for teaching “deep understanding,” as defined by the Authentic Achievement Project (Newmann & Associates, 1996). The study also found that American classrooms had many more outside interruptions than did Japanese classrooms (Stigler & Hiebert, 1999). Interruptions like announcements over the PA system and the principal or staff walking into the room are somewhat common in the United States, but very rare in Japan. The transfer to the United States of the classroom pedagogies for deep understanding might not be successful unless we also reduce the number of outside interruptions and make other cultural adjustments.

In my preliminary study of computer use in American and Japanese classrooms in the TIMSS Classroom Video database, it appears that the set up time was much longer in American classrooms than in Japanese classrooms. That is, less time was taken in getting the students started doing their work on the computer in Japan than in the United States. In addition, I noticed that American teachers were more likely to give their students trivial computer tasks—work that could be done more quickly with a calculator or an encyclopedia. As there are only 11 computer-using classrooms in the TIMSS Video database, these are not definitive findings. However, they illustrate how culture may influence the degree of productivity from technology-supported teaching activities.
Quantitative Research on Technology in Teaching • 89

The Program for International Student Assessment (PISA) is a cross-national assessment of skills and knowledge but only of students who are 15 years old. The principal subject areas are reading, mathematics, and science, with one subject predominating every 3 years as shown in Table 3.5. PISA is conducted by the OECD (Organization for Economic Development). In each testing year, each student is given 2 hours of testing plus a background questionnaire to fill out. It is important to note that as an assessment of 15 year olds, PISA does not directly correspond to most other assessments nor is it possible to generalize to students in a specific grade.

In 2000 and 2003 the background questionnaires included questions on students’ access to and their use of computers. Economists Fuchs and Woessmann (2004) analyzed the 2000 data using structural regression models predicting reading and math test performance from some of these indicators of computer access and use. They concluded that student computer use tended to have a negative effect on performance.

Fuchs and Woessmann’s argument is very weak, because, while they controlled for family-background variables, they could not identify the type or quality of computer use. Bielefeldt (2005) pointed out this and other reasons why the conclusions were a distortion of the impact that ICT has on learning and teaching. What is needed to make statements about impact are much better measures of ICT utilization including differentiation of the type and quality of the exposure to ICT. Well-designed experimental and quasi-experimental studies would greatly add to our knowledge about these types of questions.

In PISA 2006, the background questionnaire included more detailed questions about use of ICT by the students tested. Thus, the 2006 data will provide more detailed results with regard to the amount and type of ICT access, use, and capability of 15-year-olds. This type of self-report data has come to be called ICT “familiarity” as opposed to ICT-related skill or knowledge.

In 2006 PISA developed a hands-on ICT literacy assessment to run on laptops that were taken to each school where the assessment was delivered to the students on these laptops. Only a few countries opted to participate in this ICT literacy assessment and apparently the laptop administration was not only costly but problematic, as there are no plans to use this procedure in the next assessment. The report on this experience is not yet available.

<table>
<thead>
<tr>
<th>Year</th>
<th>Primary topic</th>
<th>No. of countries</th>
<th>Special features</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>Reading</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>Math</td>
<td>41</td>
<td>Problem-solving subtest</td>
</tr>
<tr>
<td>2006</td>
<td>Science</td>
<td>58 (estimated)</td>
<td>Computer-based testing option</td>
</tr>
<tr>
<td>2009</td>
<td>Reading</td>
<td>(unknown)</td>
<td>(unknown)</td>
</tr>
</tbody>
</table>
At the time of this chapter writing, OECD PISA is making a decision as to whether or not in 2009 the main assessment will be reading or ICT literacy. The original plan was that the subject would be reading and that plan will probably prevail.

Other studies. While both national and international school-based assessments of ICT knowledge and skill appear to be waning, ICT familiarity surveys of older populations are becoming more commonplace. One organization conducting many such studies is the Pew Internet & American Life Project (http://www.pewinternet.org/about.asp). Another is the World Internet Project (http://www.worldinternetproject.net/) involving 23 countries. The latter is the brainchild of Jeffrey Cole of the University of Southern California, and the project makes it possible for countries to conduct comparable surveys. Their goal is to conduct an annual survey but many, if not most, of the countries are unable to participate in any given year.

The greatest limitation of these types of studies is that all of their indicators of new literacy and use of new media are self-report measures. In that sense, they are more superficial than the studies given major attention in this chapter. Typically, their samples are taken of adult populations, which means that the results cannot be compared directly to school-age populations but sometimes allow for comparison of various adult age groups. Like opinion polling in general, the methodologies of these studies must be carefully evaluated. Whenever methodological details are impossible or difficult to obtain, it signifies that the study’s methodological foundation may be weak.

Methodological Issues

Issues pertaining to research methodology are best organized in terms of the goals that are intended and the methods selected to achieve those goals. This structure is represented in Table 3.6, which shows four goals as columns and two types of methods as rows. The cells give illustrative approaches, e.g.,

Table 3.6  The Intersection of Major Goals and Methods Used in Large-Scale Assessment-Related Research: Illustrative Approaches

<table>
<thead>
<tr>
<th>GOALS</th>
<th>Exploration</th>
<th>Theory/hypothesis generation</th>
<th>Assessment and monitoring</th>
<th>Impact studies and evaluations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methods</td>
<td>Case studies</td>
<td>Grounded theory</td>
<td>Constructed response coding</td>
<td>Formative evaluation</td>
</tr>
<tr>
<td>Qualitative Research:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observation and Interviewing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantitative Research:</td>
<td>Pilot or field testing</td>
<td>Causal theory building</td>
<td>Traditional and online assessment</td>
<td>Summative evaluation</td>
</tr>
<tr>
<td>Surveying and Assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Downloaded By: 10.3.98.93 At: 05:23 20 May 2019; For: 9781410618894, chapter3, 10.4324/9781410618894.ch3
grounded theory, that are typical of a given approach when it is used in the pursuit of a given goal category.

This table contrasts both qualitative and quantitative approaches by listing them as methods. While large-scale empirical research tends to be largely quantitative, qualitative approaches are needed for problems that are not well understood and for which indicators have not been refined. Often a combination of qualitative and quantitative approaches is needed in cross-national research due to the highly diverse cultural contexts to be studied (Creswell, 1998; Yin, 1994).

The prior section highlights the major, large-scale empirical studies that relate to new media and new literacies. In the remainder of this section, the discussion reviews the methodological decisions that are largely unique but critical to quantitatively oriented, international comparative assessments. The first methodological issue addressed is that of survey and test design, including item and test construction. Other issues addressed include sampling design, data collection, and data analysis. The discussion suggests criteria that can be used in evaluating existing studies and in planning future ones. These issues are given time and attention here because they are largely neglected in textbooks and the scholarly literature in general.

Survey and Test Design

The first step that needs to be completed after a conceptual framework is the refinement of research questions. Some core or foundational questions underlying all hunches and hypotheses should be derived to serve as orienting devices for priority setting in the project. Research questions also serve a second function of communicating the essence of the study to persons outside the project.

In the course of specifying the research questions, the key units of analysis will become clear. This is important because the units of sampling will be based on them. Most of the studies considered here use several units for sampling and analysis—student, teacher, school are the typical units selected.

In order to answer the research questions formulated, survey instruments will be needed within which are draft items for measuring the required indicators. Quite possibly these questionnaires will need to be revised for different grade levels if multiple grades are targeted. Some questionnaire instruments required for a typical study are listed in Table 3.7. Each instrument and its codebook must be prepared for each population or grade; however, often the manuals can be combined for all populations or grades.

The construction of the survey questionnaires will necessarily be guided by the indicators and constructs to be measured, and the constructs will be identified and elaborated by an analytical framework. A challenging aspect of such an analytical framework is the multilevel structure involving constructs and variables not only for students but also for teachers and schools. Once an
analytical framework is specified, attention must be devoted to operationalizing the various constructs and factors desired as variables for the analysis.

Sampling Design

The first design step after a conceptual framework and research questions is the delineation of the populations (or universes) to be investigated. From the standpoint of international, comparative assessment in education, generally this means choosing among the alternatives of selecting a grade, an age, or students with a specific number of years of schooling completed. Initially, the IEA studies selected grades but PISA selected an age, 15. Now the IEA studies use a combination of these criteria. A grade is selected by looking in each country for the grade with the most students who are age x. Alternatively it may be the grade with the most students who have completed y years of schooling. This type of population definition is necessary to standardize the bases of comparison across countries.

Whether a grade or an age is the basis for selecting sampling units, the selection is called a population. (See Table 3.6.) Often a single grade will be selected from the primary grade level and one from the lower secondary levels. These levels are called populations 1 and 2, because they will not necessarily be the same grades in each country due to differences across education systems, especially in the age that the students first start school.

### Table 3.7 Instruments and Manuals Required for a Typical Study

<table>
<thead>
<tr>
<th>Population 1 (Grade 4)</th>
<th>Population 2 (Grade 8)</th>
<th>Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Instruments</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Principal Quest.</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Tech. Quest.</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Teacher Quest.</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Student Quest.</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Performance Test</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>*<em>Manuals</em></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Design Plan</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Curriculum Analysis</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Sampling</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Administration</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Data Entry and Coding**</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Codebook</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

* Some of these manuals may be combined. The instruments (tests and questionnaires) should be included in one of these manuals or published separately.

** Data Entry and Coding Manual includes cleaning as well as entry and coding. It also should include scoring for assessment items and rules for deriving composite scores or indicators.
In SITES-M1, a target grade range was also defined for purposes of comparing school resources across educational systems where the grade groupings vary greatly by country. The concept of target grade range was needed for asking questions about the accessible infrastructure and common experiences of students in a meaningful- but-comparable grade grouping. The definition of target grade range was especially significant because not only were many questions worded to refer to these target grades, but the number of students in the grade range was used as the measure of size (MOS) for purposes of sampling and weighting. The target grade range was defined as the three grades containing the most students of age 10 and 14, respectively, for populations 1 and 2. Thus, the target grade would generally, but not always, be the middle of each of these three grades. The biggest technical difficulty with the target-grade-range specification is that in some systems a school-level boundary falls somewhere within the grade range. A guideline for such situations can be to take the school level with the two grades with the most age-eligible students and define those two grades as the target grade range. In such situations, a sampling expert is needed to help adjudicate the choice, taking both the interests of the international coordination and that of the national study into account.

With respect to population coverage, each study must set standards, and especially critical is the maximum exclusion level. In any sampling at the national level, there will be schools that are extremely difficult to include, e.g., prison schools, special-education schools, or even schools that are located in extremely remote areas. An international study will typically allow exclusion of such schools but only up to a total of 3%, 5%, or a similar threshold.

Sampling design is one of the most crucial elements in the methodological quality of these studies. Most large-scale school surveys or assessments use stratified PPS (probability proportionate to size) random sampling. When a random class or classroom is selected for an assessment, cluster sampling is being used as well. Stratified sampling reduces sampling error when the stratification variables define groups that are homogeneous with respect to the dependent variables. Sometimes the biggest challenge is to find a database of all schools with good stratification variables like school type, SES, and region. PPS requires a meaningful MOS (measure of size) variable, which in most school surveys is the size of school as measured by school enrollment. Using PPS makes it possible to adjust the probabilities of selecting any school proportionate to its size. Without such an adjustment, small schools, or students in small schools, have much too high a chance of being included in the final sample.

Nonresponse is perhaps the most difficult and challenging sampling problem. If the nonresponse rate is lower than 50%, it implies that there may be a systematic bias introduced into the sample. Studies set completion rates as standards that have to be met. For example, the minimum completion rate set for SITES-2006 was 85% for schools and 85% for teachers, which means any nonresponse rate greater than 15% is unacceptable. Some projects or
researchers set lower minimum completion rates. Others allow for replacement samples, that is, completely separate samples drawn to be used for sampling individual cases if a case from the first sample does not respond after a certain number of tries.

International studies utilize the threat of exclusion of a country’s data from publication if the minimum completion rate is not reached. Generally, a compromise is implemented. If the minimum completion percentage is not achieved, the country’s data should be flagged or placed “below the line,” which means it is placed at the bottom of the table in order to warn readers that those data are more subject to systematic or sampling error.

Sampling classrooms, teachers, and/or students randomly within schools poses another challenge. Once a principal has agreed that the school will participate, then the best procedure is to get from the principal’s office a list of classes, teachers, and/or students from which the researcher will randomly select the units to be approached for participation. Producing such lists may be demanding for the school, so some researchers attempt other procedures, e.g., the principal may be asked to distribute the questionnaires randomly to teachers. Such procedures may not produce good random samples, so they should not be used unless there is a validation study done to demonstrate that it works effectively.

**Data Collection**

Alternative modes or delivery mechanisms for the survey and assessment instruments are of major interest to researchers doing large-scale projects. While standard paper-and-pencil forms of the survey instruments may be provided, electronic-response options also should be considered when feasible and appropriate. Providing the questionnaires on a disc or CD should also be considered.

The novelty of Web-surveys already has worn off in many countries and completion rates are typically very low when questionnaires are long. Thus, multimode surveys should probably be used. The most important way to avoid bias due to electronic modes of questionnaire administration is to make the Web-delivered questionnaire as isomorphic or similar to the paper version as possible (Dillman, 2000). Other potential sources of bias include differences in respondents with respect to familiarity with and feelings toward the technologies involved. These factors need to be measured, so that they can be statistically investigated and used as controls in the analysis.

With a Web-delivery option available, small improvements in response rates should be possible. The costs of Web delivery of questionnaires in many different languages have been high in the past, but technology for this is improving. A number of technical and logistical issues need to be solved, such as developing online questionnaires that can run on most browsers and providing privacy space between students as they are taking the surveys online.
The chief challenges of using the Web for delivering performance assessments are largely technical. One of these challenges is security, which includes the maintenance of the integrity of each student’s responses as well as guaranteeing fairness by eliminating opportunities for cheating or in other ways benefiting from the work of other students. Another major challenge is understanding the comparability of Web-based performance assessments. For some purposes, it is important to be comparable to other forms of assessment. However, comparability across subjects and testing occasions is the most critical. Without such comparability, the performance assessment may give certain types of students or types of situations an unfair advantage. International projects have a number of demands that are mostly unique to that environment: translation comparability, comparability of hemispheres, scheduling conflicts with other international studies, and data quality control across countries.

The translation challenge is mainly that of ensuring that words and phrases have essentially the same meanings in each country after translation from a master document. Questionnaires and tests are the biggest translation tasks but any instructions related to the administration of the surveys must also be translated. To ensure comparability, back-translations are needed where an independent translator translates the document back to the original language. When there are disagreements, then a third party should be called in to help resolve the discrepancies. Some projects use a third-party organization to do all or most of the translation, which can be extremely expensive.

The hemispheric challenge is that schools in the northern and southern hemispheres begin their school year at different times of the year. This means that if the assessments are given at the same time of the year in both hemispheres, the schools in one hemisphere will have a half-year advantage. One issue when assessments are given at two different times is that a decision must be made as to whether or not the North goes first or the South. Since most of the countries typically participating in IEA or OECD surveys are in the North, countries in the North probably get more influence over such decisions.

Surveys and assessments require that some school staff be devoted to the study for a time, and student time becomes an issue with assessments. When there are many surveys going on in any given year, these demands become a serious burden for the schools, and it produces noticeable, if not serious, increases in nonresponse rates. When international studies are being planned, the planners always try to take the schedule of other studies using the same population(s) into account.

In order to ensure the highest possible quality control within realistic financial constraints, various structures for quality monitoring should be considered. In a large-scale, international study, a critical element will be one or more consultants hired to be data quality monitors. The data quality monitor would help to work out issues between the national and international staffs regarding the collection and processing of data. The main function of the data
quality monitor can be to facilitate communication and understanding among those involved and to work toward improving the overall data quality.

Another role within each country might be an assessment monitor. The assessment monitor would be expected to observe the assessment procedure in a random sample of schools during the administration of the assessment. Regional workshops may be offered to train the national assessment monitors. The national assessment monitors might also serve as translation judges to arbitrate among differences of opinion regarding best translations of elements in the instruments and manuals. If the procedure to be used requires two or more independent translations from English to a local language, the translation judge would be responsible for collecting and giving advice on translation options.

Data Preparation and Analysis

Technically, the most challenging aspect of data preparation is the estimation of sampling errors. In order to calculate these errors, it is necessary to calculate one or more sampling weights for each case in the sample. The weights can be calculated as a complex function of the sampling design. In addition, the weights may need to be adjusted for biases in the sample due to nonrandom missing data. A procedure called poststratification weighting is used for this class of problems to minimize the sampling bias in the data files for analysis. Any large-scale quantitative study should have the sampling expertise available to handle this aspect of the data preparation and analysis.

The analytical methods used should be guided substantively by the research questions and technically by the quality of the data. In the early phase of the project, the study directors should prepare an “analysis plan,” which should include frameworks and construct refinements, proposed indicators, and prototype tables for statistical analysis.

Before the substantive analysis begins, measurement analysis should be applied to the data in order to determine if the constructs were validated by the indicators, and if so, to what degree. To determine the quality of the assessment instruments, item response theory (IRT) analysis should be performed for each country and each population. Differential item functioning (DIF) analysis should also be applied to determine if particular items functioned differently and hence might have produced an unintended bias for a particular type of student. Concurrently with the scaling and validity tests, should be the calculation of sampling errors. Sampling-error estimates should take into account the complex sample design used by each country by using appropriate software.

To as great an extent as possible, the data should be analyzed to take into account the hierarchical structure in the data. HLM methods should be used as appropriate to test for relationships across multiple levels of the structure. This should make it possible to explore whether or not school contexts have an effect on teaching practices and orientations, and also to determine to what
extent school contexts and pedagogical orientations shape student competencies, experiences, and attitudes. Of special interest is whether the ICT competencies of the students are greater in contexts where the teachers are engaged in innovative pedagogical practices.

Conclusions and Implications for Future Research

Priorities for Deciding What to Assess

First, what are the best priorities to use as criteria for future project? Our theoretical discussion listed six continua for such consideration. Ideally, the project would have the resources to sample points along all the continua. This being unrealistic, it will be necessary to select specific levels of difficulty, levels of complexity, and explicitness in the information and knowledge to be handled, the stability (versus novelty) of the media, and the extent to which the tasks can be done without the media or technology.

Most researchers yield to the temptation to study novel media, neglecting the stable ICT applications like text editing. In educational research, the learner and productivity roles have been the context investigated for the most part. It seems like time to investigate how both students and adults are able to deal with the new literacies that are needed for fulfilling one’s role as a responsible citizen as well. Giving careful consideration of the other criteria offered as decision criteria will point out other areas that researchers should give priority to in the future.

Conceptual Frameworks

While the emphasis in this chapter has been on assessment frameworks, contextual and causal frameworks deserve a lot more attention by scholars in the new literacies field. The payoff will be especially high with causal frameworks and specific causal models. This type of conceptualizing is needed to bring more rigor into the body of advancing knowledge about new media and new literacy.

Higher Education

Ironically, although most of the large-scale empirical research on new media and new literacies has been done by researchers at universities, but none of it has been on populations in universities or the rest of higher education. Many excuses can be found for this state of affairs, but the fact remains that if such research were conducted within higher education, it could organize itself more effectively.

There always has been more funding available for ICT-related research within the primary and secondary sectors than in the tertiary system. However, now the funding appears to be drying up for the primary and secondary sectors as well. At least this is clearly the case in the United States, and no large-scale funding sources outside the United States have volunteered to fill the gap left when the U.S. sources have pulled back. Neither the IEA nor the
OECD has been able to raise the funds needed to do the ICT-related assessments proposed by either organization during the early part of the 21st century. Even the much less costly surveys of schools and teachers related to new media and new literacies are at risk for adequate funding in the near future.

**Global Vision**

Besides learning from other educational systems, international studies of innovative or exemplary schools and teaching practices offer the possibility of an overall, global view. It may be that every country’s educational system is so different that little can be generalized. However, it is more likely that a common, international vision is emerging on how information technology can best be incorporated into teaching and learning. Actually, there may be several international visions emerging, but we do not know much about these leading-edge commonalities. Unless we know such things, we cannot plan very well for the future, which inevitably will be more global in terms of economic, communication, social, and educational systems.

Because of complexities and costs, very little research on the role of technology in education is conducted comparatively or internationally. However, when sustainability and transferability of innovative, high-impact learning practices are significant objectives, then an international context is critical to understanding the essential ingredients. Cultural traditions such as parental participation in students’ computer-based homework, and national policies such as teacher technology certification, may be essential to the success of an educational intervention and have profound implications for its maintenance and transfer. This applies to transport of exemplary practices both within and across national boundaries.

**New Technologies and New Media**

An analysis of the newly emerging technologies reveals some startling ICTs and new media that will inevitably radically change social relations and raise major new social and ethical issues (Anderson, 2001). In this concluding consideration, three types of new technology (RFID, paintable computers, and artificial cells) are evaluated in terms of their implications for new literacies.

**Finding everything with RFIDs.** RFID stands for Radio Frequency Identification and consists of an electronic tagging technology that allows an object, place, or person to be automatically identified at a distance. Tags come in various sizes and shapes and can contain antennas, GPS chips, computer chips, and memory chips. Already many items in stores possess these tags. Tags can keep track of location histories, which raises privacy concerns because tags from stores could be sending back information from one’s household to all kinds of companies and persons.
The future will likely bring us inexpensive, personal RFID systems to help us keep track of every object in an office or a home. Think of keeping track of everything in your home from your desktop computer. How much easier it would be to write chapters like this! You and everything relevant to you will exist as nodes in many overlapping networks. Personal networking in concert with object networking will create the need for new literacies in much the same way that the Internet meant that new literacies were needed to deal with it.

**Wearable, paintable computers.** Wearable computers started with clumsy head-gear but have evolved into nearly invisible components of sunglasses and regular glasses wherein small computer monitors are mounted. Printable/paintable computers will help make personal networking a reality because using an ink-jet printer you will be able to print an RFID (computer) tag on every document you view on your computer.

Eventually, more complex computer chips might be painted on specific surfaces. Artists can be creative both in making chip designs and then in painting them on surfaces with other computer chips. In the same way that multimedia literacy emerged as a new literacy in the age of color computer screens, paintings of malleable computer components will require a new dimension of aesthetic appreciation as well as language for creating flexible, powerful, practical, and even beautiful computer logic configurations.

**Genetic engineering of “human” cells.** Increasingly, computer scientists are predicting that genetic engineering in concert with nanotechnology and advanced robotics in 10 to 20 years will be producing molecular processors and memory devices the size of human cells. One implication of this is that it should be possible to implant these cells in a corner of the brain, interfacing them to our natural, biological brain cells. Once that is perfected, it should be relatively straightforward to build wireless interfaces to the brain such that software routines and data are downloaded to the brain. Furthermore, it would seem that brain cells could be copied and uploaded to the Internet and then shared with one’s friends or colleagues, for truly virtual meetings, among other applications.

To build in safeguards for a person to protect his or her “self” probably will require a set of interaction scripts that are unlike any we use today. The uniqueness would be a consequence of the fact that the self consists not only of thought patterns but also of emotional patterns and spiritual routines that are built up over time from a long history of thoughts and emotions. One’s spiritual side might be thought of as network of computer logic boards that combine thoughts and feelings around one’s deeper beliefs about what is most meaningful in life.

As we come to understand our personal emotional and spiritual life better, we almost certainly will realize that its portrayal and communication will
require even newer multimedia than we have now. Consequently, new literacies will be needed to participate in these advances.

Implications for survival of the species. While it is not possible to promise that any of these future scenarios will emerge during our lifetime, we can be sure that if they do, there will be major moral and ethic issues that will have to be confronted. In fact, the survival of humanity will depend on how well we do that. Today’s teachers have a grave responsibility: the character education of tomorrow’s human beings. To get a healthy glimpse of what it will take to get the next generation ready for the era of even newer media and literacies, we would do well to project ourselves into the future and look back to the present to see what needs to be done to ensure that the human race survives.

References
Brummelhuis, A. C. A. T., (1993, April 13). *Large scale assessment of word processing skills, results from an international option to the core instruments of the COMPED-study*. Paper presented at the Annual Meeting of the AERA, Atlanta, GA.


