Year 1 Technology Studies:
Implications for Technology in Assessment

CSE Technical Report 459

Gregory K. W. K. Chung and Eva L. Baker

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
University of California, Los Angeles

December 1997

Center for the Study of Evaluation
National Center for Research on Evaluation,
Standards, and Student Testing
Graduate School of Education & Information Studies
University of California, Los Angeles
Los Angeles, CA 90095-1522
(310) 206-1532

The work reported herein was supported under the Educational Research and Development Center Program, PR/Award Number R305B60002, as administered by the Office of Educational Research and Improvement, U.S. Department of Education.

YEAR 1 TECHNOLOGY STUDIES:
IMPLICATIONS FOR TECHNOLOGY IN ASSESSMENT

Gregory K. W. K. Chung and Eva L. Baker
CRESST/University of California, Los Angeles

The major focus of this report is to document CRESST's 1996-1997 technology initiatives in two broad areas: (a) using technology to improve the quality, utility, and feasibility of existing measures, and (b) using technology to design and develop new assessments and measurement approaches available through no other means. We review current activities, evaluate the outlook for particular assessment technologies, and report on lessons learned. Implications for assessments are also discussed.

Table 1 lists the major activities during the review period. Most of the work has focused on designing and developing a computer-based architecture for an integrated assessment system. Briefly, the CRESST integrated simulation is an approach to assessment that incorporates computer- and paper-based assessments of students who engage in complex, constructed-response tasks. These tasks are based in a real-world context, and assessment occurs throughout the task and covers one or more components of the CRESST model of learning (Baker, Abedi, Linn, & Niemi, 1996). The CRESST integrated simulation (O'Neil, 1997a) and CRESST model of learning are discussed in greater detail in the next section.

Our second major activity was to review automated approaches to scoring constructed-response tasks. We identified candidate approaches to computer-based scoring of essays (Chung & O'Neil, 1997) and paper-based concept maps (O'Neil & Klein, 1997). Dissemination activities have been mainly conference participation and technology demonstrations. The remaining set of activities focused on basic research underlying cognitive issues and concept planning of how CRESST assessment technology could support school-level and district-level planning and distance learning.

Table 1

List of Major Activities, Project 1.3, Technology in Action, 1996-1997

Area	Activity	Status
CRESST integrated simulation	Design and implement the integrated simulation assessment architecture.	Completed
	Design and implement a Java version of the individual concept mapper.	Completed
	Design and implement a HyperCard-based collaborative concept mapper.	Completed
	Design and implement Web search-based problem-solving tasks and measures.	Completed
Enhancing scoring and feasibility of performance assessments	Review methodological approaches to the automated scoring of essays.	Completed
	Feasibility study of machine scoring of concept maps.	Completed
	Computer-based assessment of problem solving.	Completed
	Concept planning for next-generation computer-based performance assessments.	Completed
Dissemination	Present CRESST work at conferences.	Ongoing
	Provide training for CRESST-developed assessment tools.	Ongoing
	Conduct technology demonstrations of CRESST assessment tools.	Ongoing
Related activities	Preliminary design for Quality School Portfolios.	In progress
	Conduct basic research on concept mapping transfer.	In progress
	Conduct basic research on cognitive processing with conceptual models.	In progress
	Preliminary design for stand-alone negotiation simulation.	Completed
	Concept planning for integrating CRESST assessments into distance learning environments.	Completed

CRESST Integrated Assessment Simulation

CRESST cognitive model of learning. All assessment technology development has been guided by CRESST's model of learning. The model broadly characterizes learning as a function of content understanding, problem solving, self-regulation, collaboration, and communication. Table 2 lists brief definitions of each component. See Baker (1995), Baker et al. (1996), and Klein, O'Neil, Dennis, and Baker (1997) for a detailed discussion.

Table 2
CRESST Model of Learning Components

Component	Definition
Content understanding	Understanding of subject matter content, which includes domain concepts, facts, principles, and procedures
Problem solving	Activity directed at attaining a goal when the solution is not obvious. Problem solving involves content understanding, problem solving strategies, metacognition, and motivation.
Collaboration/teamwork	Working with other members of a team to jointly complete a task.
Self-regulation	Includes metacognition, effort, self-efficacy.
Communication	The ability to express oneself clearly and effectively for various audiences and purposes.

The CRESST model has provided the theoretical context for the design and development of several assessment technologies, and has led to the conceptualization of technology as an important component in the effort to measure complex student performance. Our assessment approach has been two-pronged. First, we employ a suite of assessment tools (rather than a single, monolithic tool) to measure specific components of the CRESST model. Second, we integrate the assessment tools with the task structure to provide a problem-based, relatively authentic context for students to work in. We refer to this as an *integrated simulation* approach to assessment. Assessment is integrated in that students are assessed on each component of the CRESST model one or more times as they go through the task. By simulation we mean approximating in a computer environment a real-world context for students.

The implementation of the integrated simulation in a computer-based environment provides new assessment opportunities to measure student performance on complex, constructed-response tasks (Bennett, 1993). A computer-based environment affords opportunities to measure complex student performance not feasible in any other environment. This point is essential and provides a clear and compelling rationale for the use of technology in the measurement of complex performance. Unlike traditional performance assessments, which are based on widely varying data sources and implementations, our integrated simulation approach provides a relatively stable measurement context (thus reducing methodological concerns), provides an open-ended environment that can vary in task complexity (thus providing opportunities for students to demonstrate a range of skills), and provides an extraordinarily rich environment to measure not only the products of student performance, but also the process of student learning. Evidence of student cognitive processes is typically limited to one-shot measures (e.g., self-reports on questionnaires) or time- and labor-intensive methods (e.g., analyzing think-aloud protocols or behavioral observations). In our integrated simulation, we can continuously measure more directly what students are doing as they engage in a range of cognitively demanding tasks. Furthermore, these measurements are unobtrusive and inexpensive. With the CRESST model of learning and the integrated simulation as the context, we next describe our integrated simulation system.

CRESST Integrated Assessment Simulation System

Background. During 1996-1997 CRESST began in earnest to design and implement an integrated simulation. Our goal is to integrate task, technology, and assessment to provide multiple content areas for multiple grades and audiences, and to simulate an environment that provides an authentic, real-world, problem-based context. Ideally, students using our integrated simulation would be required to demonstrate a range of cognitive skills to deal effectively with the complexity and "messiness" of the environment. The conceptualization of an integrated simulation is the result of years of programmatic research (e.g., Baker, Gearhart, & Herman, 1994; Baker & Niemi, 1991; Baker, Niemi, & Herl, 1994; Baker, Niemi, Novak, & Herl, 1992) about how to best integrate existing CRESST research with technology to produce an assessment platform that would be far more economical than domain-dependent, highly customized assessment measures.

Our approach to the design of the integrated simulation was to focus on both assessment and technology issues. Starting with the CRESST research base and experience (paper- and computer-based), we asked ourselves what assessment measures (a) could be adapted to a computer-based, integrated simulation environment with a reasonable chance of success, (b) would result in an order-of-magnitude increase in utility or value, and (c) would provide valid measurement options that would not exist otherwise. In addition to these assessment issues, we evaluated technology issues such as (a) the maturation of client/server technology, (b) the long-term outlook of an Internet/Web presence in educational settings, (c) the changing relationship between technology costs and capability, and (d) the availability of development tools to reduce software development costs.

These assessment and technology issues led to both the adoption of existing CRESST measures and the development of new, technology-based assessments. The existing measures include use of short-answer responses to measure prior knowledge and essays to measure content understanding (Baker, Aschbacher, Niemi, & Sato, 1992) and use of a questionnaire to measure self-regulation (O'Neil & Abedi, 1996). These measures are included in part to provide a traditional "feel" to some of the assessments. The new measures include networked computers to measure teamwork (O'Neil, Chung, & Brown, 1997), and Internet/Web-based problem-solving measures linked to search behavior and search performance (Bates, 1989; Borgman, Hirsh, Walter, & Gallagher, 1995; Moore, 1995; Schacter et al., 1997). The result was a loosely coupled system comprised of well-established and understood paper-based assessments, computer versions of paper-and-pencil measures, and new computer-based assessments. The CRESST integrated simulation was adopted by the Computer-Aided Education and Training Initiative (CAETI) project (see Herl, O'Neil, et al., 1996). Herl, O'Neil, et al. tailored the design and measures to meet specific CAETI program requirements. Table 3 lists the major integrated simulation activities.

System Architecture

Background. A long-term design goal of the CRESST integrated simulation is to create a unified system around an Internet/Web-based client-sever

Table 3

CRESST Integrated Simulation Activities, 1996-1997

Activity	Description
Design integrated simulation system architecture	Activity directed at achieving a domain-independent architecture for a client-server, Internet/Web-based system. Architecture designed to support multiple content areas, multiple measurement opportunities, and range of task complexities.
Design and develop individual concept mapper in Java	Activity directed at developing a Java-based concept mapper. Java is a write-once, run-anywhere language. Used existing HyperCard concept mapper as the design model.
Design and develop search/ problem-solving task and measures	Activities directed at developing measures for measuring problem solving based on search behavior and search performance in the integrated simulation.
Design and implement HyperCard-based collaborative concept mapper	Activities directed at the collaborative version of the HyperCard-based concept mapper.

architecture. Such an architecture makes it easier to support scalability and extensibility. By scaleable we mean creating a system to handle 1 to n users with minimal performance degradation, and by extensible we mean the capability to add content or assessment tools as needed with minimal cost and system impact.

Our initial version implements the basic architecture with two assessment tools: a concept mapper and a search/problem solver. Figure 1 shows the architecture of the current system. In this configuration, the client (i.e., student) accesses the Web site, which contains domain-specific information, such as texts on environmental science. A concept mapper is available for use while searching through the information. Student concept maps can be scored in real-time by the software, and feedback is provided to the student almost instantaneously. In addition, an explicit link between the concept map and the information space is established by a bookmarking feature—requiring students to "bookmark" Web pages they found relevant to a particular concept in the concept map. Part of the flexibility of the architecture is that the task defines how the tools (i.e., concept mapper, bookmarking, information space) are used. Figure 2 shows an example screen shot of the Web interface.

As an example of the application of integrated simulation, Herl, O'Neil, et al. (1996) had students do the following. Students first created a concept map on environmental science. This map was based on their existing knowledge of the subject. Students during this phase did not have access to any additional information. After completing their initial concept map, the maps were scored and

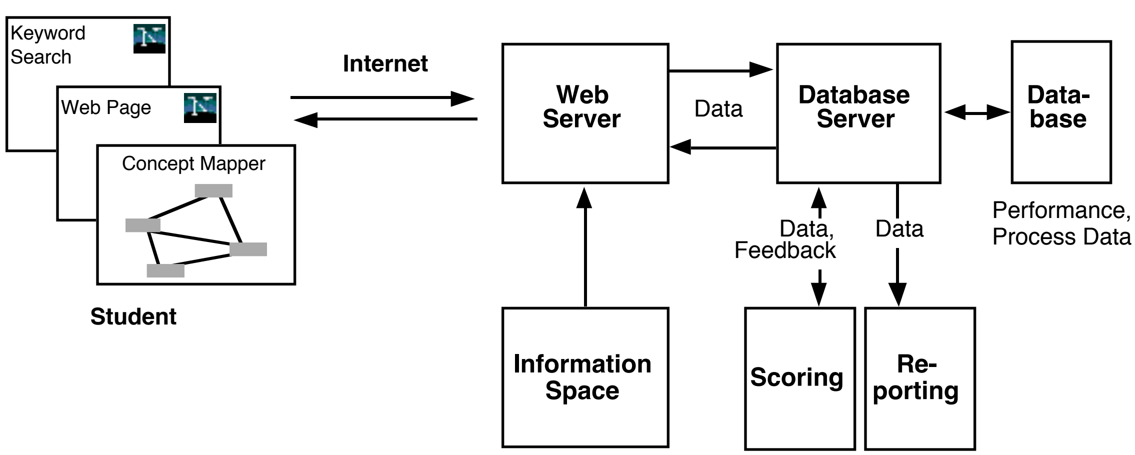

Figure 1. Integrated simulation architecture.

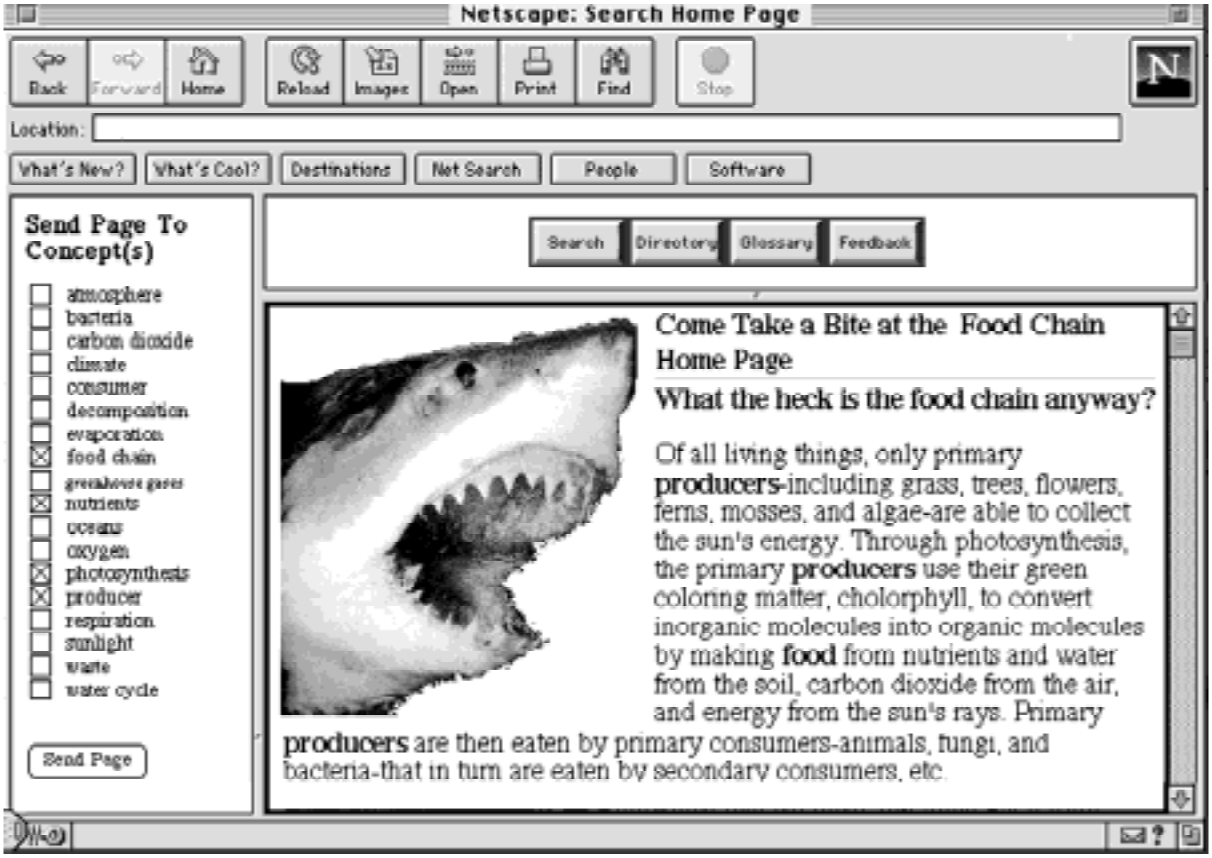

Figure 2. Example Web page that contains the concepts food chain, nutrients, photosynthesis, and producer. Bookmarking is handled by checking the appropriate concepts on the left side of the screen and clicking on the "Send Page" button.

general feedback returned to the students about which concepts "needed work." At this point, students had access to the information space—Web pages on environmental science. Students could search for information, modify their concept maps, and request feedback. The one activity that was requested of students was for them to bookmark their concepts. That is, when students found information they believed relevant to a concept in their map, they were told explicitly to bookmark that page.

Web-based information space. A major component of the integrated simulation is a Web information space. This space is domain-specific (i.e., specific to particular content area), and contains information relevant to the task. By relevant we mean that the Web pages were related in some way to the content. In our initial version we omitted nonrelated Web pages primarily because we believed that the inclusion of nonrelevant pages would create a daunting search task. From a technical standpoint, having a Web site as an information source provides a simple and flexible way to provide content-specific information. Creating new content means adding Web pages to the information space. Just as important, by establishing the content base online we can monitor students' access to the information. Depending on the task configuration, questions—such as where students went in the information space, how long they spent on particular pages, what information students were searching for, and how much time was spent on relevant versus less relevant information—can all be answered via analyses of server Web page access logs. Such questions could not be answered cost-effectively using noncomputer-based approaches.

An important aspect of our integrated simulation is that the information space is self-contained (i.e., students cannot leave our Web site). This feature constrains the information students use, so we can link students' use of particular information with their performance on, for example, a concept mapping task. As an example of the utility of a constrained information space, the CAETI program (Herl, O'Neil, et al., 1996) used a concept mapping task in conjunction with a search task. Herl et al. rated each Web page relative to each concept used in the concept mapping task. Students were asked to "bookmark" pages they judged to be relevant to a concept. This bookmarking feature linked students' relevancy judgment of a particular Web page to a particular concept. Thus, Herl et al. could examine the relationship between the students' relevancy judgments, the quality of information students accessed, and the quality of the students' concept maps, providing measures of how well students were able to find and use good information to improve their concept maps.

Keyword search. Another capability designed into the integrated simulation is a simple keyword search facility providing AND and OR Boolean operators. Students can search the site for information by typing in search terms, and use the Boolean operators to limit or expand their searches. The search terms and operators are logged by the Web server. By logging what students typed in for their search, we can derive measures of search performance such as the use of search terms relative to the task, the sophistication of search use (i.e., the use of Boolean operators while searching), and ultimately the structure of the search strategy.

SQL database. One critical component of the architecture is the database. We use a database server that supports SQL (structured query language), an industry-standard database language. The database allows us to maintain student performance data over time. As an example of the utility of a database, in the CRESST integrated simulation system, students can create, save, and retrieve concept maps at any time and from any Internet accessible computer. Without a database, students would not be able to store any information. Thus such information would not be available for assessment purposes.

Real-time scoring. One advanced feature of our system is that we have implemented real-time scoring for concept maps. Student concept maps can be scored on demand. The concept map configuration is downloaded to the server, and the server software compares the student map against a set of expert maps. Feedback is returned to the student in a form that lists the concepts that need "a lot of work," "some work," and "little work." Herl, Baker, and Niemi (1996) describe the scoring algorithm and approach in detail. We are pursuing real-time computer scoring for performance on other elements of the assessment model.

Assessment technology outlook. The outlook on using Web-based technology to deliver online assessments remains promising. A convergence of different factors makes Web-based assessments timely. These factors include the maturation of the technology, increasing availability and affordability of the Internet to home and education markets, and federal support for Internet access for schools. The Web has reached critical mass. CRESST's experience with Web-based technology for *assessment purposes* is unique and timely.

Integrated simulation lessons learned:

- **Transaction model of processing required.** CRESST is using Web technology for purposes quite different from typical Web applications. Most Web sites are read-only. We are using the Web site in a read-write mode—a transaction model more akin to real-time multi-user database systems (e.g., automated teller machines). Thus, concurrency issues (dealing with simultaneous transactions) have been a continuous concern with our system.

- **Scalability is an issue.** The (assessment) use of the Web server is such that multiple users (e.g., a classroom of students) are using the system simultaneously. While this is not an issue for simply accessing Web pages, the situation changes when concurrent users perform computationally intensive transactions such as saving a concept map or requesting feedback on their concept maps. Our current system can support up to 15 simultaneous students. Additional hardware will be needed to support more students.

- **Assessment model is essential.** The design of a Web-based assessment system is no different from designing other systems, particularly in the need to have a framework to work in. The CRESST model of learning and integrated simulation provided a framework to think about how to leverage technology to measure student learning.

Individual Concept Mapper

Background. Over the last year we developed a Java version of an individual concept mapper. A concept map is a node-link-node representation of content, where nodes represent concepts and links represent relationships between connected concepts (Dansereau, 1995; Jonassen, Beissner, & Yacci, 1993). Figure 3 shows the Macintosh-based, HyperCard-based concept mapper, and Figure 4 shows the Java-based concept mapper. The rationale for developing a Java version was threefold. First, the Java version would provide a concept mapping tool that could be used across different computer platforms. Java is a

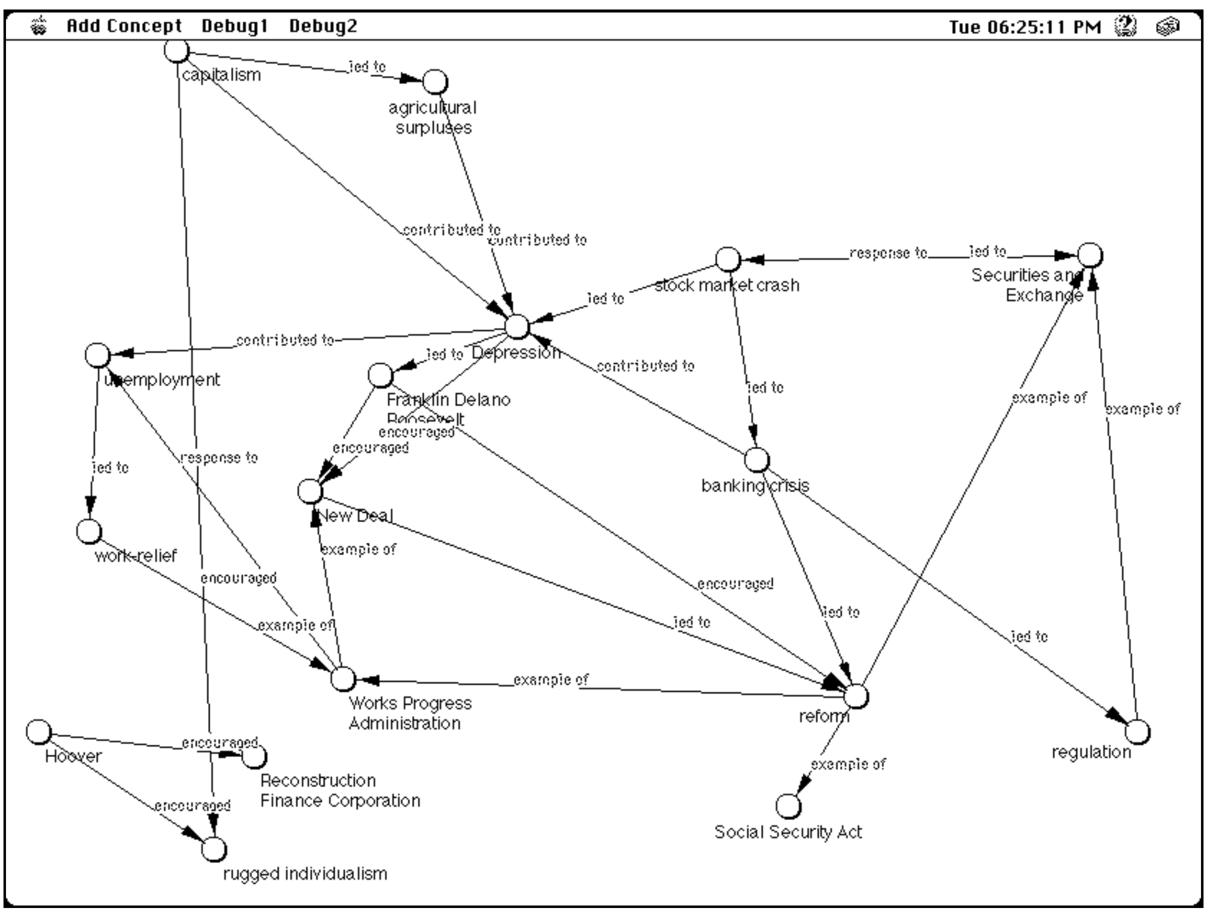

Figure 3. HyperCard-based individual concept mapper.

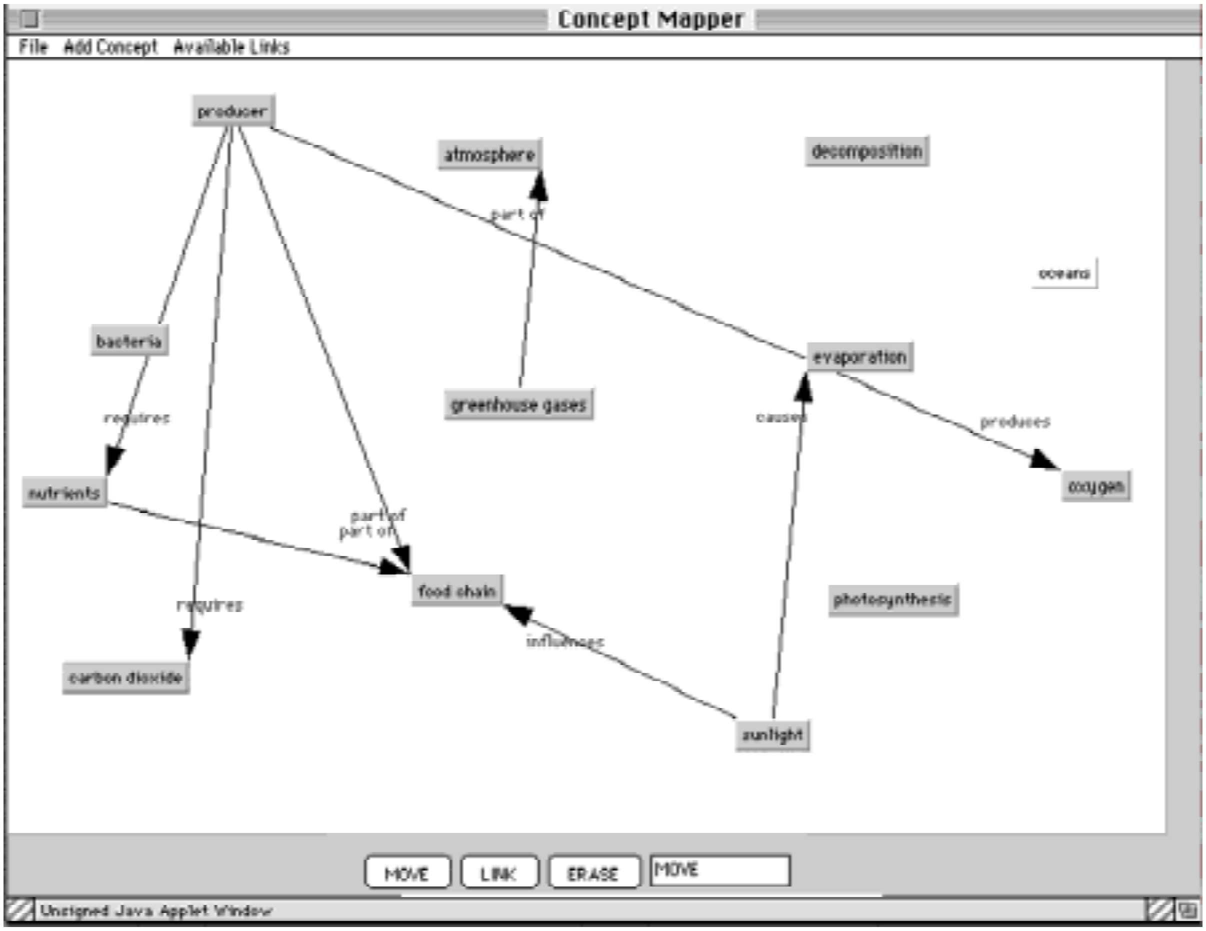

Figure 4. Java-based individual concept mapper.

platform-independent language supported on all major operating systems and Web browsers. The second reason is that given our Internet/Web-based architecture, a Java-based, Internet deployable concept mapper fits well into our long-term goal of having an integrated suite of assessment tools. Given the experience with concept mapping (i.e., existing HyperCard concept mapper, existing research base on paper-based concept mapping, and existing in-house expertise), this task was a logical, low-risk/high-payoff first step. Table 4 shows the major differences between the HyperCard and Java versions of the individual concept mapper.

A typical concept mapping task consists of providing the student with a fixed set of concepts and links. A student is instructed to construct a map of his or her understanding of how the given concepts relate to each other. Students are free to configure their maps any way they choose, and they can add, delete, or move

Table 4

Summary of Differences Between the Hypercard- and Java-Based Individual Concept Mappers

Feature	HyperCard-based concept mapper	Java-based concept mapper
Stand-alone version	Yes. Able to run concept mapper on any Macintosh.	Partial. Able to create a concept map but no capability to save a map to the local machine. This is a security constraint of Java applets.
Internet deployable	No.	Yes. Users can create, save, and reload concept maps to and from a server.
Real-time scoring	Yes.	Yes.
Platform	Macintosh.	Macintosh, Windows, and UNIX. Any operating system that supports Java.
Authorable	Partial. Users able to specify concepts and links through ASCII text files.	No.
Ease-of-use	Uses Macintosh-specific user interface elements. Program "feels" like a typical Macintosh application.	Uses Java standard interface elements. Application conforms to "lowest common denominator" usage. Application, relative to HyperCard version, is slightly more cumbersome to use.

concepts and links at will. The rationale for using concept maps in assessment is that they are constructed-response tasks that measure content understanding (Herl, Baker, & Niemi, 1996).

Assessment technology outlook. The Java-based concept mapper is a clear example of what we think is the future of online assessment software. The tool is easily learned and requires less than 10 minutes of training (Herl, O'Neil, et al., 1996). Our successful deployment of the concept mapper over the Internet to both Macintosh and Windows platforms provides reassurance in Java technology.

Individual concept mapper lessons learned:

- **Existing design model facilitates development.** The development of the Java-based concept mapper was facilitated by the existence of a HyperCard version of the mapper. The HyperCard version provided a working design model that laid out much of what was expected in terms of functionality and user-interface.

- **Java is still coding.** Java, like any other programming language, means coding. Despite Java being a modern object-oriented language, there is still a substantial amount of programming that must be done to produce a product like the concept mapper. The real benefit of Java is that it has become an industry standard, supported by all major operating systems, thus providing an unprecedented level of multi-platform support.

- **Generality vs. functionality trade-off.** One drawback of Java is that it is a language of the least-common denominator with respect to the user-interface. Because of the requirement to support different user-interface elements across different operating systems (e.g., the use of one, two, or three mouse buttons), Java has been designed for the most general case. Much of the richness of a particular operating system is inaccessible.

- **User interface needs improvements.** Our initial effort was directed at creating a functional Java concept mapper with less focus on the visual aspects of the interface. However, we recognize that the concept mapper could be improved to provide a more seamless operation (e.g., in Figure 4, replacing the "Move," "Link," and "Erase" buttons with a more intuitive design). In addition, the concept mapper "look" needs to be enhanced visually (e.g., more attractive node and link displays, or the inclusion of graphics instead of boxes for nodes).

- **Apparent acceptance by teachers and students.** Our informal discussions with teachers and students suggest that they enjoy using the concept mapper and view the task as a valuable one. Teachers see concept mapping as being performance oriented, and students are genuinely engaged with on-line concept mapping.

Automated Data Logging

Background. The third component of the integrated simulation architecture is data logging. By data logging we mean the capture and storage of students' data while they are using the integrated simulation (e.g., concept map state or Web pages accessed). For this component, our short-term approach was to rely on the Web server logging (i.e., the server log and the access log) and to develop custom software to extract the log information where necessary. This design decision was driven by two factors: (a) lack of programming resources during 1996-1997 to implement a robust logging and reporting system, and (b) anticipation that lessons learned from CAETI would help clarify requirements for a fully automated data logging system. Thus, during the first year we assumed a mix of manual and automated data processing.

Server log. In general, a Web server records all activities related to system activity, status, and database transactions. Our interest was in the latter information. The server log contains all keyword searches performed by users. From the server log we can derive measures of keyword searching.

Access log. The access log contains a history of all Web pages accessed. Each access log entry contains the following information: IP address of the client machine, the date and time of the access, the fully qualified URL of the Web page, additional information such as page size, request status, and referring URL. For our purposes, we were only interested in the IP address, date and time, and URL. The IP address provides a way for us to identify individual computers (and thus users), and the URL provides us with the necessary information to derive measures of search behavior (e.g., browsing).

Concept map data tables. Students' concept maps are saved in database tables. Each time a student saves a map, the concept map information is saved as a separate entry. Thus, we have available different states of progress for students' maps. While currently not used for analyses, the opportunity exists to analyze students' concept maps over time.

Assessment technology outlook. While we did not commit a lot of resources on the automated data logging component during 1996-1997, we continue to believe that automated data logging remains an essential part of any Web-based assessment system. One outstanding design issue is whether to rely on the server-supplied data logs or to develop a custom data logging facility.

Collaborative Concept Mapper

Background. During 1996-1997 we designed and developed a collaborative version of the concept mapper for the Macintosh. The collaborative mapper was programmed in HyperCard and is not Java-based. We implemented the collaborative mapper in HyperCard because we were confident we could develop a HyperCard version given our experience (e.g., O'Neil et al., 1997). We believed a Java version would have been too risky an effort given no in-house Java experience. Thus, the collaborative mapper is a stand-alone component of the integrated simulation. The development of a Java-based collaborative mapper is planned.

To integrate collaborative services into the concept mapper we used the built-in networking capabilities of the Macintosh operating system and HyperCard. Figure 5 shows a sample screen shot of the collaborative mapper. A typical task is to assign a group of three students to jointly construct a concept map. The members of the group are connected via a network and are assigned anonymous identifiers (i.e., "M1," "M2," or "M3"). One member of the group is

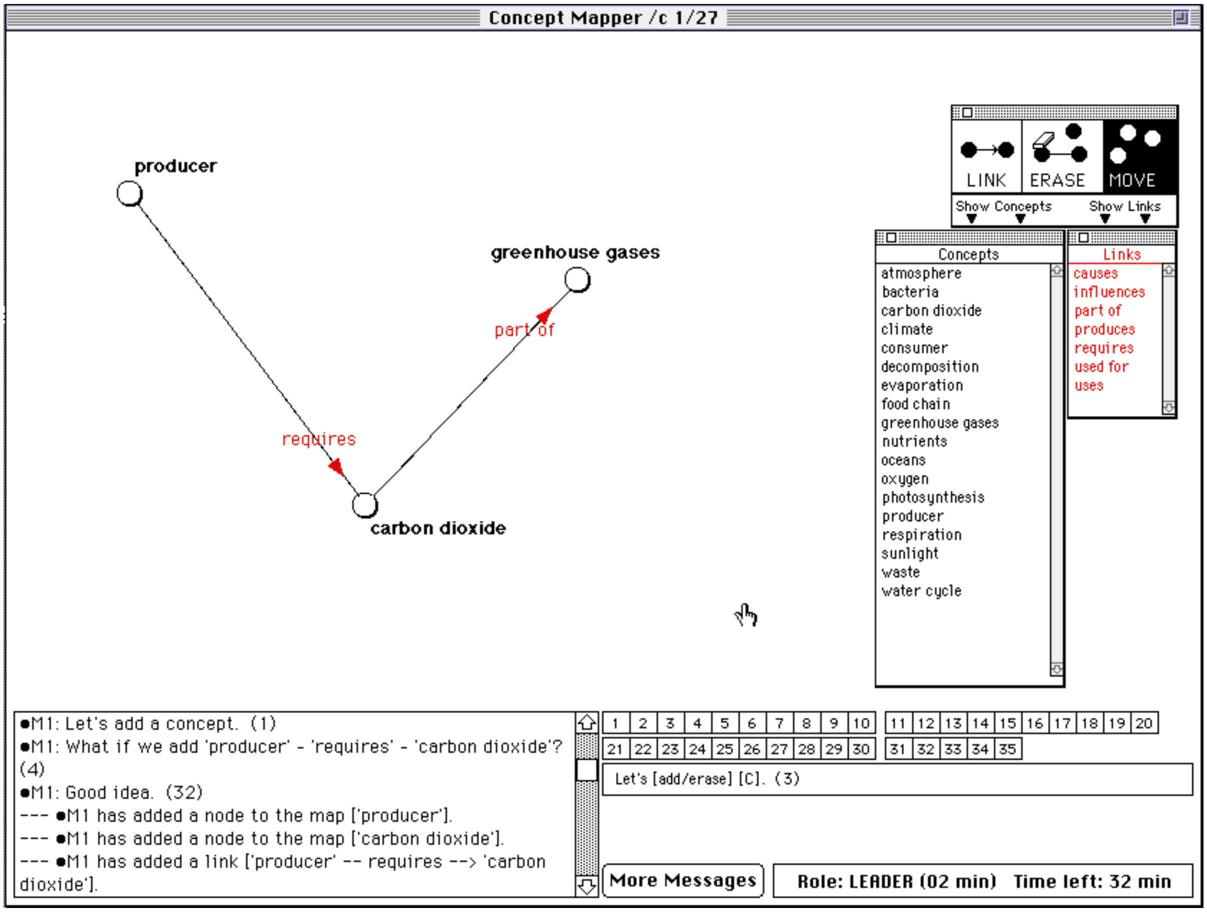

Figure 5. Collaborative concept mapper screen shot.

initially assigned the role of the leader. Leadership rotates throughout the task, and only the leader can change the map. Nonleaders are instructed to advise the leader on a course of action. All computers are updated as changes occur (e.g., someone sending a message, or the leader making changes to the concept map); thus, the computers are synchronized with each other. Communicating between group members is done through the use of pre-defined messages. Members are given a list of 37 messages, and they send these messages to each other (e.g., "Let's link *carbon dioxide* to *producer*."). The rationale for using predefined messages is to provide a means to measure team processes in real-time. All messages are coded a priori as reflecting a particular team process. The message coding scheme is based on the work of O'Neil et al. (1997).

Assessment technology outlook. Our work over the last three years with some form of a networked-based, collaborative task has yielded generally positive results (Herl, O'Neil, et al., 1996; O'Neil et al., 1997). The technology is generally not a problem, students enjoy using the system, and groups can complete the task with predefined messages. We think that our general approach to using networked computers as a means to set up a collaborative environment remains viable. An outstanding issue is determining the best way to measure teamwork. Our current approach is to provide students with predefined messages, and to consider the quantity and type of message sent as an index of different teamwork processes. Our long-term goal is to develop a Java/Web version of the collaborative mapper, and integrate this version into our existing Web server architecture.

Collaborative mapper lessons learned:

- **Interface for predefined messages is problematic.** One finding during usability testing is that students find the use of predefined messages difficult. The messages are hard to use and at times do not express what users want to communicate. We are in the process of refining our message set.

- **Students can complete the task.** Despite the difficulty of the predefined messages, students are able to use the messages and complete the task. What is unknown is whether the students completed the task because of or in spite of the messages.

- **Enjoyable and engaging.** Students find the collaborative map activity fun and engaging. Communicating with other people over computer networks is novel and fun for students.

Enhancing Scoring and Feasibility of Performance Assessments

Another set of activities have involved examining the feasibility of machine scoring of essays and paper-based concept maps. Table 5 briefly lists the activities.

Automated scoring of essays. During 1996-1997, we began the groundwork for the analysis of techniques associated with automated scoring of essays. A review of the field turned up two candidate approaches: Project Essay Grade (Page & Petersen, 1995) and latent semantic analysis (Landauer & Dumais, 1997). A report of these two approaches is given in Chung and O'Neil (1997). Strengths and weaknesses, assessment potential, and long-term outlook for automated scoring of essays is covered in this report.

Table 5
Enhancing Scoring and Feasibility of Performance Activities, 1996-1997

Activity	Description
Automated scoring of essays	Examine different approaches to automated scoring of essays. Assess feasibility and utility.
Automated scoring of concept maps	Examine different approaches to automated scoring of paper-based concept maps. Assess feasibility and utility.
Computer-based assessment of problem solving	Develop conceptual approach to measuring problem solving within the framework of the CRESST model of learning. Suggest measurement approaches in computer-based environment.
Advanced software for computer-based assessments	Develop high-level specifications for computer-based assessment tools designed to measure various components of the CRESST model of learning. Specifications included needed hardware, software, and level-of-effort.

Automated scoring of concept maps. O'Neil and Klein (1997) conducted a feasibility study on machine scoring of paper-based concept maps. Two approaches were considered. The first approach was forms-based. Working with a test form designer (e.g., National Computer System), a concept map form would be developed. Students are provided with a list of concepts and links. Students are then asked to select the most important terms and begin connecting these terms using the appropriate links. Concepts in the form are labeled with letters; links are labeled with numbers. A legend at the bottom of the form lists the concepts and links and the associated letter or number. The attractiveness of this feature is that the forms are easy to score and turn-around time is quick.

The second approach is to allow free-form drawing of concept maps, and handle the data entry with voice-recognition technology. In this scenario, students would draw a concept map from scratch. The concepts and links would be provided to students, but unlike the concept map form, students can draw, erase, and connect the concepts in any order. The only constraint would be that the students include an identifying letter (for concepts) or number (for links) on their maps.

The data entry would then be a matter of reading a three-character letter-number-letter sequence. Because voice recognition technology can be trained to recognize single letters and numbers with high accuracy, this approach provides a rapid way of entering node-link-node information. Once entered into the computer, the same scoring software used in the computer-based concept mapper can be used to score the maps.

Computer-based assessment of problem solving. O'Neil and Schacter (1997) have reviewed the literature on problem solving and developed specifications for the measurement of problem solving in computer-based environments. Based on the CRESST model of learning, O'Neil and Schacter see problem-solving as being comprised of four elements: (a) content understanding, (b) problem-solving strategy use, (c) metacognition, and (d) motivation. O'Neil and Schacter suggest measuring content understanding and problem-solving strategy use domain specifically (i.e., measures are based on the specific content and task), and to measure metacognition and motivation domain independently. Some issues raised are the need for a conceptual framework, what to measure, assessment task and format, purpose of testing (e.g., program evaluation or diagnostic), unit of analysis (e.g., individual or teams), testing time, and consequences (i.e., high or low stakes).

Advanced software for computer-based assessment. In another set of activities we evaluated potential next steps for assessment software. Drawing on CRESST experience and lessons learned on various technology projects (e.g., Baker, Niemi, & Herl, 1994; Baker, Niemi, et al., 1992; Herl, O'Neil, et al., 1996; O'Neil, 1996; O'Neil, Allred, & Dennis, 1992; O'Neil et al., 1997), preliminary high-level specifications were developed for assessment tools that would measure one or more components of the CRESST model of learning. In general, Chung, Klein, Herl, and Schacter (1997) and Chung, Klein, Herl, O'Neil, and Schacter (1997) identified two layers of software necessary to support a diverse suite of assessment tools. First, application program interfaces must be developed to provide a reusable set of software components. These components would provide common functions and services to the assessment tools, maximizing the amount of software reuse. Second, the assessment application itself should operate independently or jointly. The set of assessment tools outlined in Chung, Klein, Herl, O'Neil, & Schacter (1997) include individual and team-based simulations, text processing applications, procedural mappers, multimedia concept mappers, problem-solving tools, and information organizers such as outliners and idea generators. Collectively, these tools would provide (a) state-of-the-art constructed-response tasks, (b) opportunities for performance-based assessment with respect to the CRESST model of learning, (c) measurement opportunities only available through computer-based means, and (d) substantive learning opportunities for students. The feasibility of this expanded tool development is now under review.

Dissemination Activities

See the Appendix for a complete citation of all dissemination activities. Table 6 lists the different kinds of dissemination outlets for 1996-1997.

Related Activities

Table 7 lists an additional set of project activities that have important implications for future work.

Table 6

Dissemination Activities, 1996-1997

Activity	Description
Conferences	Publicize CRESST integrated simulation and research conducted in that framework.
Training	Training activities related to the use of the concept mapper and integrated simulation.
Assessment technology demonstrations	Demonstrations of computer-based assessment tools.

Table 7

Project 1.3 Related Activities, 1996-1997

Activity	Description
Quality School Portfolios	High-level design of a system to enhance school- and district-level planning and monitoring.
Concept mapping transfer	Investigate transfer effects of learning concept mapping in multiple content areas.
Cognitive processing with conceptual models	Investigate cognitive processing with conceptual models.
Stand-alone negotiation simulation	Design requirements for simulating team members based on O'Neil et al.'s (1997) union-management simulation.
Distance learning assessment	Concept planning for the integration of CRESST assessment tools into a planned distance learning program.
Certification testing	Use of concept mapping as a technique to certify job performance.

Quality School Portfolio (QSP). Over the last year, this activity has focused on rapid prototyping of demonstrations and different user interfaces. Redesign work began in the latter half of 1996-1997 and included preliminary content and technical specifications to keep up with changes in the computer platform and software environments. These specifications were developed to address schools' Title I reporting needs. Preliminary content specifications covered desired content for an initial version of QSP. Preliminary technical specifications included descriptions of screens, functions associated with the screens, and flowcharts of interrelationships between screens. A working prototype is being refined for trial use in local schools and to assist in Title I data collection.

Concept mapping transfer. This activity focused on investigating transfer effects using concept maps. Klein (1997) investigated the effects of students learning concept mapping in one or two subjects, with and without metacognitive self-monitoring training. Klein hypothesized that students who both engaged in self-monitoring and were exposed to two subject areas would form better schemata, engage in greater metacognitive activity, and perform better on the transfer measure than other students. Some support was found for the beneficial effects of monitoring on schema formation. In addition, even with a relatively brief treatment period, at-risk students were able to learn the cognitive strategy of concept mapping, to engage in metacognitive activities such as self-monitoring, to construct good concept mapping schemata, and to transfer to a large degree.

Cognitive processing with conceptual models. This activity focused on investigating the cognitive processes learners invoke while using visual conceptual models to understand expository text. Visual conceptual models are iconic representations of concepts, which are related to one another by arrows or physical proximity, and collectively represent a cause-effect system. This research will provide detailed accounts of cognitive processing in relation to problem solving and retention.

Stand-alone negotiation simulation. One short activity was to investigate how the union-management software (O'Neil et al., 1997) could be modified to simulate the presence of missing team members. This would allow the use of the software by one person while the behavior of the other two team members would be simulated. The proposed solution was to use a model-based approach to simulate the sending of messages from other team members.

Distance learning assessment. This activity was directed at developing a preliminary plan of CRESST's role in the joint UCLA/USC distance learning program in multimedia (O'Neil, 1997b). In this plan, CRESST would provide the assessments (both traditional and performance-based). Computer-based assessments (e.g., the CRESST integrated simulation) are expected to be part of the assessment package and will likely include collaborative and individual components.

Implications for Assessment

Our approach at the outset was to use technology in ways that went beyond simply mimicking paper versions of assessments. We assumed that a far better use of technology would be to leverage the unique capabilities of technology to provide a clear advantage in terms of cost, utility, validity, reliability, access, or accommodation. Several major themes have emerged from our experience over the last year: (a) Technology affords unique measurement opportunities; (b) technology initiatives hinge on software development capability; (c) operational planning is important for long-term success; and (d) measurement issues remain unchanged (e.g., need for reliability and validity).

Computer-based assessments provide the capability to measure complex learning. One of the most promising aspects of assessment technology is the capability to have students engage in constructed-response tasks and to measure both student performance and student learning processes. This capability is one of the most compelling reasons for using technology in assessment. Our experience to date points to the feasibility of developing powerful assessment environments that will provide authentic challenges to students. While this idea is not new and underlies many performance assessments, what is new—and only technology can feasibly provide this—is the capability to measure unobtrusively and more completely students' learning as they learn. The leverage computer-based assessments provide is the capability to design in measurement points virtually at will and at any point in the interaction between student and computer. As an example, an on-going dissertation (Dennis, 1997) is studying the dynamic modeling of some learning uses of concept maps.

However, despite having the capability, the placement of measurement points in the task must be driven by a cognitive model of student learning. For example, in the CRESST integrated simulation students are required to bookmark pages they believe to be relevant to particular concepts. Our assumption is that bookmarking requires students to evaluate the material on the Web page and make a judgment about the relevancy of the information. Bookmarking is just one example of a measurement point. Other examples of measurement points used in CRESST software are listed in Table 8. Useful application of the idea of measurement points will occur only when the task is closely integrated with the human-computer interface of the assessment system. The challenge is to design a task and interface that require students to interact with the computer. Ideally, this interaction reflects the results of students' cognition. We think that capturing behavior that reflects complex student thinking as students carry out the task will provide a far more complete picture of student performance.

Technology initiatives hinge on software development capability. One clear outcome of our experience over the last year related to software development capability. The goal can be accomplished by creating in-house capability or by out-sourcing. From an assessment standpoint, in-house software development means being able to tailor the technology to meet very specific assessment needs and can result in an order-of-magnitude increase in value. An example of what we consider a high-payoff project is the Java-based concept

Table 8

Example of Measurement Points in the CRESST Integrated Simulation

Measurement point	Description
Bookmark	A trace of students' bookmarks. Bookmarks provide a measure of what students considered relevant and can be compared to relevancy judgments of experts.
Search terms used	A trace of students' use of search terms is useful for measuring sophistication of search strategies (e.g., use vs. non-use of Boolean operators).
Web page access	A trace of Web page access reveals access patterns of students. Can also reveal navigation pattern of students throughout the information space.
Concept map state	A snapshot of students' concept maps at intervals throughout the task may reveal students' growth of understanding over time.
Concept and links events	Data on when concepts and links are created, deleted, or modified may reveal how students went about constructing their concept maps.
Predefined messages	Data on the particular message sent by each student in the collaborative concept mapper. Also reveals the kind of message sent (i.e., what teamwork process category the message belongs to) and how much communication occurred among students.

mapper. The mapper is easily used, scored, and deployable across different machines over the Internet.

On the other hand, software development is complicated, time consuming, and very risky. Large-scale software development is one of the most complicated processes in the world. Software development is not simply "writing code." A successful software project is the result of having clear requirements about functionality, tools to support design, development and testing, competent people to translate functional requirements to design specifications and code, and knowledgeable project management that can provide support and direction. Changing requirements result in continuous changes in design, coding, and testing. Poor design can result in complicated code that is neither maintainable nor extensible. Inadequate tools can result in time-consuming manual testing and debugging.

Software development capability becomes increasingly important as the scope of the assessment and measurement needs grow. Given the unique work done at CRESST, there are no off-the-shelf products that can be used to assess students in the way we want. There are products that can be used to mimic different parts of the CRESST integrated simulation (e.g., Inspiration [1997] for concept mapping), but they have no capability to measure student performance and processes. We are experimenting with both in-house and out-sourcing approaches.

Operational planning is important for long-term success. As assessment needs grow and deployment expands beyond research needs, the need for operational planning become increasingly important. By operational "smarts" we mean the knowledge, experience, and know-how to design and deliver robust, scaleable, and extensible systems. Software that has "real" end-users with "real" needs requires operational capability. Such capabilities include (a) dedicated, high-performance hardware that is continuously online 24 hours a day, 7 days a week, (b) controlled access to the system requiring a user accounting system, (c) fault tolerant hardware so that the system can recover gracefully from hardware failures, and (d) daily system backups to archive data. These requirements also mean upgrading systems on a more routine basis.

Measurement issues remain unchanged. Although much of 1996-1997 activities have been devoted to issues of feasibility, utility, and cost, we recognize the importance of validating our systems not as technology systems, but as assessment systems that deliver high-quality measurement. Issues of validity and reliability become increasingly important and more complex as new assessment formats go online. For example, although there has been work on validating paper-based concept mapping (e.g., Herl, Baker, & Niemi, 1996), we have just begun to gather data on student performance on online concept maps (e.g., Herl, O'Neil, et al., 1996). We are only beginning to understand the relationship between task, online behaviors and processes, student performance on concept maps and searches, and the usefulness of different measures toward characterizing student performance in an online environment. If online assessments are to be taken seriously as alternative forms of performance assessment, future work must be directed at addressing the reliability and validity of online assessments.

Future Activities

One of our major activities over the next year will be to address the assessment issues of our online assessments. We plan to conduct validity and reliability studies on these assessments in 1997, including their use for assessing students with special needs. Studies are planned that will explore concept mapping in non- or limited-English contexts (e.g., Korean language students or English language learners). We also expect to continue to integrate different assessment tools into the CRESST integrated simulation, and to incorporate an authoring shell for our concept mapper so that various users (e.g., teachers) can specify their own concepts and links for our concept mapper. A third major activity is to continue to develop the prototype of the Quality School Portfolio. Finally, we expect our dissemination outlets to continue to be major education conferences and technology demonstrations.

REFERENCES

Baker, E. L. (1995, September). *Finding our way*. Presentation at the annual conference of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California, Los Angeles.

Baker, E. L., Abedi, J., Linn, R. L., & Niemi, D. (1996). Dimensionality and generalizability of domain-independent performance assessments. *Journal of Educational Research, 89*, 197-205.

Baker, E. L., Aschbacher, P. R., Niemi, D., & Sato, E. (1992). *CRESST performance assessment model*s. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Baker, E. L., Gearhart, M., & Herman, J. L. (1994). Evaluating the Apple Classrooms of Tomorrow. In E. L. Baker & H. F. O'Neil, Jr., (Eds.), *Technology assessment in education and training* (pp. 173-197). Hillsdale, NJ: Lawrence Erlbaum Associates.

Baker, E. L., & Niemi, D. (1991). *Assessing deep understanding of science and history through hypertext*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Baker, E. L., Niemi, D, & Herl, H. (1994). Using HyperCard technology to measure understanding. In E. L. Baker & H. F. O'Neil, Jr., (Eds.), *Technology assessment in education and training* (pp. 133-152). Hillsdale, NJ: Lawrence Erlbaum Associates.

Baker, E. L., Niemi, D., Novak, J., & Herl, H. (1992). Hypertext as a strategy for teaching and assessing knowledge representation. In S. Dijkstra (Ed.), *Instructional models in computer-based learning environments* (pp. 365-384). Berlin: Springer-Verlag.

Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review, 13*, 407-424.

Bennett, R. E. (1993). On the meaning of constructed response. In R. E. Bennett & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing, and portfolio assessment* (pp. 1-27). Hillsdale, NJ: Lawrence Erlbaum Associates.

Borgman, C. L., Hirsh, S. G., Walter, V. A., & Gallagher, A. L. (1995). Children's searching behavior on browsing and keyword online catalogues: The science library catalogue project. *Journal of the American Society for Information Science, 46*, 663-684.

Chung, G. K. W. K., Klein, D. C. D., Herl, H. E., O'Neil, H. F., Jr., & Schacter, J. (1997). *Scaling up CRESST assessment software* (Deliverable to OERI, Award No. R305B60002). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Chung, G. K. W. K., Klein, D. C. D., Herl, H. E., & Schacter, J. (1997, February). *Computer-based automated assessment tools suite: A proposal*. Internal correspondence, University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Chung, G. K. W. K., & O'Neil, H. F., Jr. (1997). *Methodological approaches to on-line scoring of essays* (Deliverable to OERI, Award No. R305B60002). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Dansereau, D. F. (1995). Derived structural schemas and the transfer of knowledge. In A. McKeough, J. Lupart, & A. Marini (Eds.), *Teaching for transfer: Fostering generalization in learning* (pp. 93-121). Mahwah, NJ: Lawrence Erlbaum Associates.

Dennis, R. A. (1997). *Growth modeling of the effects of performance feedback task and scoring procedure on concept mapping performance*. Manuscript in preparation, University of California, Los Angeles, Graduate School of Education & Information Studies.

Herl, H. E., Baker, E. L., & Niemi, D. (1996). Construct validation of an approach to modeling cognitive structure of U.S. history knowledge. *Journal of Educational Research, 89*, 206-218.

Herl, H. E., O'Neil, H. F., Jr., Dennis, R. A., Chung, G. K. W. K, Klein, D. C. D., Lee, J., Schacter, J., & Baker, E. L. (1996). *Technical report of year 1 CAETI findings* (Report to ISX/DODEA). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Inspiration 4.0 [Computer software]. (1997). Portland, OR: Inspiration Software.

Jonassen, D. H., Beissner, K., & Yacci, M. (1993). *Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Klein, D. C. D. (1997). *I've seen this before? The effects of self-monitoring and multiple context instruction on knowledge representation and transfer among middle school students*. Unpublished doctoral dissertation, University of California, Los Angeles.

Klein, D. C. D., O'Neil, H. F., Jr., Dennis, R. A., & Baker, E. L. (1997, March). The five families of cognitive learning: A context in which to conduct cognitive demands analyses of innovative technologies. In H. F. O'Neil, Jr. (Chair), *An integrated simulation approach to assessment*. Symposium presented at the annual meeting of the American Educational Research Association, Chicago.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211-240.

Moore, P. (1995). Information problem solving: A wider view of library skills. *Contemporary Educational Psychology, 20*, 1-31.

O'Neil, H. F., Jr. (1996, September). *Lessons from CRESST technology research programs*. Presentation at the 1996 Annual Conference of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles.

O'Neil, H. F., Jr. (Chair). (1997a, March). *An integrated simulation approach to assessment*. Symposium presented at the annual meeting of the American Educational Research Association, Chicago.

O'Neil, H. F., Jr. (1997b). USC/UCLA Master's of science in multimedia via distance learning. Concept paper. Unpublished manuscript, University of California, Los Angeles/University of Southern California.

O'Neil, H. F., Jr., & Abedi, J. (1996). Reliability and validity of a state metacognitive inventory: Potential for alternative assessment. *Journal of Educational Research, 89*, 234-245.

O'Neil, H. F., Jr., Allred, K., & Dennis, R. A. (1992). *Simulation as a performance assessment technique for the interpersonal skill of negotiation* (Deliverable to OERI, Contract No. R117G10027). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

O'Neil, H. F., Jr., Chung, G., & Brown, R. (1997). Use of networked simulations as a context to measure team competencies. In H. F. O'Neil, Jr. (Ed.), *Workforce readiness: Competencies and assessment* (pp. 411-452). Mahwah, NJ: Lawrence Erlbaum Associates.

O'Neil, H. F., Jr., & Klein, D. C. D. (1997). *Feasibility of machine scoring of concept maps* (Deliverable to OERI, Award No. R305B60002). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

O'Neil, H. F., Jr., & Schacter, J. (1997). *Test specifications for problem solving assessment* (Deliverable to OERI, Award No. R305B60002). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Page, E. B., & Petersen, N. S. (1995). The computer moves into essay grading: Updating the ancient test. *Phi Delta Kappan, 76*, 561-565.

Schacter, J., Herl, H. E., Chung, G. K. W. K., O'Neil, H. F., Jr., Dennis, R. A., & Lee, J. J. (1997, March). Feasibility of a Web-based assessment of problem solving. In H. F. O'Neil, Jr. (Chair), *An integrated simulation approach to assessment*. Symposium presented at the annual meeting of the American Educational Research Association, Chicago.

APPENDIX

DISSEMINATION ACTIVITIES

Books

Baker, E. L. (Ed.). (in press). *The third generation of educational performance assessments*. Mahwah, NJ: Lawrence Erlbaum Associates.

O'Neil, H. F., Jr. (Ed.). (1997). *Workforce readiness: Competencies and assessment*. Mahwah, NJ: Lawrence Erlbaum Associates.

Chapters

Baker, E. L., Herman, J. L., & Gearhart, M. (1996). Does technology work in schools? Why evaluation cannot tell the full story. In C. Fisher, D. C. Dwyer, & K. Yocam (Eds.), *Education and technology: Reflections on computing in classrooms* (pp. 185-202). San Francisco: Jossey-Bass.

O'Neil, H. F., Jr., Allred, K., & Baker, E. L. (1997). Design of teacher-scored measures for workforce readiness competencies. In H. F. O'Neil, Jr. (Ed.), *Workforce readiness: Competencies and assessment* (pp. 177-204). Mahwah, NJ: Lawrence Erlbaum Associates.

O'Neil, H. F., Jr., Allred, K., & Baker, E. L. (1997). Review of workforce readiness theoretical frameworks. In H. F. O'Neil, Jr. (Ed.), *Workforce readiness: Competencies and assessment* (pp. 3-26). Mahwah, NJ: Lawrence Erlbaum Associates.

O'Neil, H. F., Jr., Allred, K., & Dennis, R. A. (1997). Use of computer simulation for assessing the interpersonal skill of negotiation. In H. F. O'Neil, Jr. (Ed.), *Workforce readiness: Competencies and assessment* (pp. 205-228). Mahwah, NJ: Lawrence Erlbaum Associates.

O'Neil, H. F., Jr., Allred, K., & Dennis, R. A. (1997). Validation of a computer simulation for assessment of interpersonal skills. In H. F. O'Neil, Jr. (Ed.), *Workforce readiness: Competencies and assessment* (pp. 229-254). Mahwah, NJ: Lawrence Erlbaum Associates.

O'Neil, H. F., Jr., & Baker, E. L. (1997). A technology-based authoring system for assessment. In S. Dijkstra & N. Seel (Eds.), *Instructional design: International perspectives. Vol. II: Solving instructional design problems* (pp. 113-133). Mahwah, NJ: Lawrence Erlbaum Associates.

O'Neil, H. F., Jr., Chung, G. K. W. K., & Brown, R. S. (1997). Use of networked simulations as a context to measure team competencies. In H. F. O'Neil, Jr. (Ed.), *Workforce readiness: Competencies and assessment* (pp. 441-454). Mahwah, NJ: Lawrence Erlbaum Associates.

Journal Articles

O'Neil, H. F., Jr., & Abedi, J. (1996). Reliability and validity of a state metacognitive inventory: Potential for alternative assessment. *Journal of Educational Research, 89,* 234-245.

O'Neil, H. F., Jr., & Brown, R. S. (in press). Differential effects of question formats in math assessment on metacognition and effect. *Applied Measurement in Education*.

O'Neil, H. F., Jr., Sugrue, B., & Baker, E. L. (1996). Effects of motivational interventions on the National Assessment of Educational Progress mathematics performance. *Educational Assessment, 3,* 135-157.

Conferences

Baker, E. L. (1996, March). *Technical developments in performance assessment*. Presentation at the Center for Educational Technology, Florida State University, Tallahassee.

Baker, E. L. (1996, April). *Reflections on education and technology, 1985-1995*. Paper presented at the annual meeting of the American Educational Research Association, New York.

Baker, E. L. (1996, May). *Technology*. Presentation at the 1996 Milken Family Foundation National Education Conference, Los Angeles.

Baker, E. L. (1996, June). *Why? Evaluation drivers including cost, time and training and blockers including access, equity and classroom practices*. Presentation at the Forum on the Application of Emerging Technologies to Student Evaluation and Assessment: What? Why? How? and When?, Los Angeles.

Baker, E. L. (1996, October). *New Developments in Assessment*. Presentation at the 1996 Annual Conference of the Association of California School Administrators, San Francisco.

Chung, G. K. W. K., Baker, E. L., Dennis, R. A., Herl, H. E., Huang, K. L., Klein, D. C. D., Lee, J. J., Schacter, J. (1996, November). *An Internet-based on-line assessment system*. Presentation at the annual meeting of the California Educational Research Association, Marina del Rey, CA.

Chung, G. K. W. K., O'Neil, H. F., Jr., Herl, H. E., & Dennis, R. A. (1997, March). Use of networked collaborative concept mapping to measure team processes and team outcomes. In H. F. O'Neil, Jr. (Chair), *An integrated simulation approach to assessment*. Symposium presented at the annual meeting of the American Educational Research Association, Chicago.

Herl, H. E., O'Neil, H. F., Jr., Chung, G. K. W. K., & Dennis, R. A. (1997, March). Feasibility of an on-line concept mapping construction and scoring system. In

H. F. O'Neil, Jr. (Chair), *An integrated simulation approach to assessment*. Symposium presented at the annual meeting of the American Educational Research Association, Chicago.

Huang, D., & O'Neil, H. F., Jr. (1997, March). *The role of parental expectation, effort, and self-efficacy in the achievement of high- and low-track high school students in Taiwan*. Poster presentation at the annual meeting of the American Educational Research Association, Chicago.

Klein, D. C. D., O'Neil, H. F., Jr., Dennis, R. A., & Baker, E. L. (1997, March). The five families of cognitive learning: A context in which to conduct cognitive demands analyses of innovative technologies. In H. F. O'Neil, Jr. (Chair), *An integrated simulation approach to assessment*. Symposium presented at the annual meeting of the American Educational Research Association, Chicago.

Mulkey, J. R., & O'Neil, H. F., Jr. (1997, March). *Effects of high-stakes computer-based examinations on self-efficacy and worry*. Roundtable presentation at the annual meeting of the American Educational Research Association, Chicago.

O'Neil, H. F., Jr. (1996, September). *Lessons from CRESST technology research programs*. Presentation at the 1996 annual conference of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles.

O'Neil, H. F., Jr. (1997, February). *Collaborative concept maps*. Presentation to the Board on Testing and Assessment, National Academy of Sciences, National Research Council, Washington, DC.

O'Neil, H. F., Jr. (1997, March). *Assessment issues in distance learning*. Presentation to the Army Science Board, Arlington, VA.

O'Neil, H. F., Jr. (Chair). (1997, March). *An integrated simulation approach to assessment*. Symposium presented at the annual meeting of the American Educational Research Association, Chicago.

O'Neil, H. F., Jr. (1997, March). Discussant. In L. T. Curran (Chair), *Testing for dollars: Effects of motivation*. Symposium presented at the annual meeting of the American Educational Research Association, Chicago.

O'Neil, H. F., Jr. (1997, March). *Technology-based assessment in the Department of Defense*. Presentation to the Board on Testing and Assessment, National Academy of Sciences, Orlando, FL.

O'Neil, H. F., Jr. (1997, May). *Problem-solving assessment*. Presentation at the OECD/International Life Skills Survey meeting, Paris.

O'Neil, H. F., Jr. (1997, May). *Teamwork assessment*. Presentation at the OECD/International Life Skills Survey meeting, Paris.

Schacter, J., Herl, H. E., Chung, G. K. W. K., O'Neil, H. F., Jr., Dennis, R. A., & Lee, J. J. (1997, March). Feasibility of a Web-based assessment of problem solving. In H. F. O'Neil, Jr. (Chair), *An integrated simulation approach to assessment*. Symposium presented at the annual meeting of the American Educational Research Association, Chicago.

Stevens, R. (1996, September). *Model-based computer assessment of problem-solving skills in science*. Presentation at the 1996 annual conference of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Angeles.

Training

Dennis, R. A., & Herl, H. E. (1996, February). Demonstration of group and individual collaborative mapper to community of educational technology developers.

Assessment Technology Demonstrations

Baker, E. L., O'Neil, H. F., Jr., Herl, H. E., & Chung, G. K. W. K. (1996, December). *Collaborative concept mapping and integrated simulation*. Demonstration to an anonymous foundation.

Baker, E. L., O'Neil, H. F., Jr., Herl, H. E., & Chung, G. K. W. K. (1997, January). *Collaborative Concept Mapping And Internet-Based Assessment System*. Demonstration to the Undersecretary of Education.

Baker, E. L., O'Neil, H. F., Jr., Herl, H. E., & Chung, G. K. W. K. (1997, February). *Collaborative concept mapping and internet-based assessment system*. Demonstration to Hughes Aircraft.

Baker, E. L., O'Neil, H. F., Jr., & Schacter, J. (1997, January). *Collaborative Concept mapping and internet-based assessment system*. Demonstration to Hewlett-Packard.

Chung, G. K. W. K., Baker, E. L., Dennis, R. A., Herl, H. E., Huang, K. L., Klein, D. C. D., Lee, J. J., & Schacter, J. (1997, January). *An Internet-based assessment system*. Demonstration to visiting professors from the Republic of China.