The Design and Development of an Evaluation System for Online Instruction

Tim L. Wentling
Scott D. Johnson

University of Illinois at Urbana-Champaign

With the growth in popularity of online instruction has also come the concern for accountability and the need to make adjustments and improvements in online programs. This paper describes the conceptualization and development of an evaluation system that can be used to judge online instructional efforts. The evaluation system addresses concerns of both program administrators and course instructors. Computer technology is used to provide partial automation to reduce respondent burden and to efficiently use evaluation resources.

Keywords: Program Evaluation, Evaluation System, Online Instruction

Online or web-based learning is an education and training phenomena that is growing rapidly around the globe. As with any innovation, there are questions about effectiveness, efficiency, and utility (McCollum, 1998). With online instruction, the issue of evaluation and accreditation is especially important due to various threats commonly perceived by educators and trainers. These threats include a change in student campus life and the roles of professors (McCollum, 1998), the fear that the push for distance learning comes from entrepreneurs and university "bean counters" rather than educators (Mendels, 1999), and whether virtual students feel more isolated than their traditional counterparts (Arenson, 1998). These concerns can only be addressed through systematic research and evaluation efforts.

Human resource development professionals are blessed with a rich history of evaluation thought, development, and research, primarily derived from the fields of education (Alkin, 1969; Patton, 1997; Scriven, 1967; Stake, 1967, 1978, 1981; Stufflebeam, 1967, 1985; Stufflebeam, Foley, Gaphart, Guba, Hammond, Merriman, & Provus, 1971; 1997) and human resource development (Holton, 1995; Kirkpatrick, 1976; 1996; 1998; Kaufman & Keller, 1994). From these foundations, specialized evaluation processes and systems have been designed and implemented for HRD interventions, including training programs (Raab, Swanson, Wentling, & Clark, 1992, Wentling, 1980, 1992, 1993).

The design of evaluation systems for online instruction has been attempted by a number of professionals, including instructional designers of CBT systems (Clark, 1994; Draper, 1996; Simonson, 1997) technology experts (Jackson, 1990; Kimball, 1998; Middleton, 1997), and HRD practitioners (Magalhaes & Schiel, Pisik, 1997). The work of these individuals demonstrates progress in applying intuitive principles and practices of evaluation to online environments. Other authors have provided practical applications of evaluation theory by offering suggestions and guidelines for the evaluation of online instruction (Khan, 1997; Nichols, 1997; Oakes, 1997; Ravitz, 1997; Thorpe, 1993).

Problem Statement

The evaluation of online instruction is an important part of the design and implementation process. Even though practitioners and academics currently evaluate online instruction, their attempts have been limited to the use of traditional research methods and intuitive approaches to evaluation. There appears to be a lack of systematic evaluation of online programs that is built on evaluation theory and practice. The activity described in this paper is an attempt to conceptualize and develop a transferable, adaptive evaluation system for online instruction.

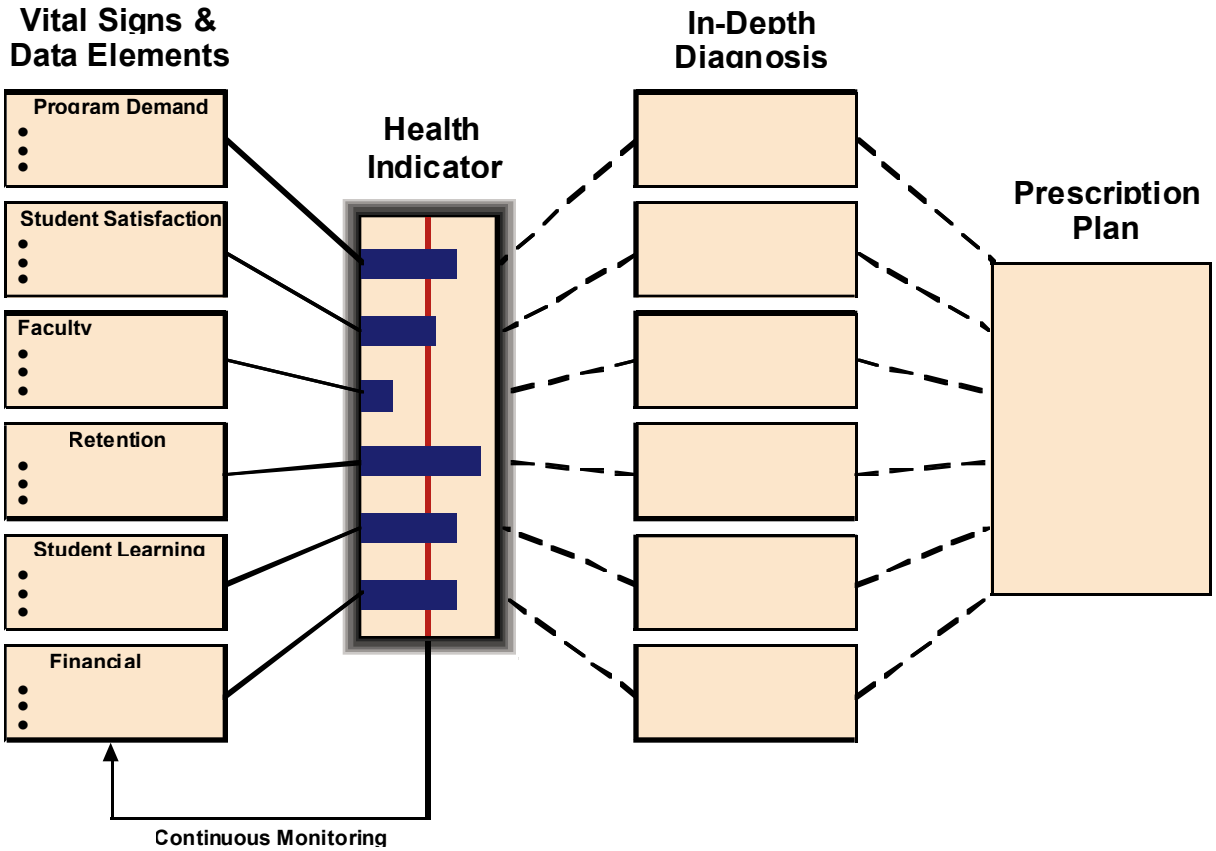
Conceptualization and Development of the Evaluation System

The purpose of the project was to design, develop, and implement an evaluation system that would meet the needs of the developers and sponsors of online instructional programs. The activities of the project reflect major efforts to obtain information, design components of the evaluation system, and test and revise the evaluation system. The specific activities of the project were to (1) develop a conceptual model for online evaluation, (2) identify specific vital signs of an online program and determine appropriate measures, (3) automate the collection and analysis of evaluation data through knowledge engineering approaches, (4) create an electronic performance support system to assist in program evaluation, and (5) conduct a field test of the evaluation system. The following sections describe each of the five major activities that are being accomplished in the formulation of the Illinois Online Evaluation System. At the time of this writing the evaluation system is evolving. Many of the tasks are complete, while some are underway.

Activity 1: Development of a Conceptual Model for Online Evaluation

This activity involved the conceptualization and design of the overall evaluation system, with the identification of major components and their functional relationships. The evaluation system, as currently designed, occurs in three stages: (1) vital sign assessment, (2) in-depth analysis, and (3) program improvement planning (see Figure 1). The evaluation system can best be explained using a medical analogy in which a physician examines a patient's vital signs to determine the patient's current state of health. Vital signs that are below acceptable standards are examined in more detail, utilizing more precise information and investigative techniques. As a result of the analysis, a plan is developed to facilitate improvement.

Figure 1. The Online Evaluation System.



Vital sign assessment (Stage 1) diagnoses the general health of the online program and individual online courses using data collected through routine activities. Example data include the number of inquiries about the program, the number of applications received, and the performance of students in the individual courses. The data are used to calculate a program's "health" rating in six areas: (1) student demand, (2) student retention, (3) student satisfaction, (4) faculty satisfaction, (5) student achievement, and (6) financial efficiency. The ratings help program personnel and sponsors monitor the overall performance of the program and individual courses to identify areas that may be problematic.

During the *in-depth analysis* (Stage 2), any vital signs that are identified as substandard are analyzed in detail. A thorough investigation is conducted to determine underlying problems or causes. Often this investigation requires the collection of additional data, or the use of data from related vital signs. For example, a low rating in student satisfaction might lead to the further analysis of the individual items that comprise the student rating indicator. Additionally, a review of data related to student learning and faculty satisfaction might be done.

Program improvement planning (Stage 3) provides solutions to the problems that are investigated during the in-depth analysis. Alternative strategies for solving the problems are identified, along with resource requirements for implementing the solutions. Program personnel are able to select various courses of action and develop plans for addressing the identified problems.

Activity 2: Vital Sign Identification and Development of Measurement Instruments and Procedures

The major activity in the development of the evaluation system involved several inquiry endeavors to identify "quality indicators" for use as assessment criteria. Lists of quality indicators drawn from the literature on computer mediated education, outcomes of education, and evaluation were identified and prioritized. An initial list of 18 "vital signs," developed from the results of the literature was reviewed by a group consisting of HRD and evaluation experts. The list of vital signs was reduced to six by determining the relative importance of each to the stakeholders of the evaluation.

Instruments and procedures for gathering, summarizing, and analyzing vital sign data are being developed by project staff. Data used to calculate vital sign ratings are obtained through analysis of electronic archives, document review, surveys, student testing, and expert review. Sample data elements used to calculate each vital sign rating are described in Table 1. The actual vital sign ratings are calculated from student test scores, mean scores on survey instruments (i.e., student satisfaction and faculty satisfaction), and data comparisons (e.g., enrollment, retention, financial data). Standard transformations of the data are performed to provide a common scale for each vital sign to facilitate portrayal and comparison across the vital signs.

Table 1
Data Elements for Vital Sign Rating Calculation.

Vital Signs	Data Elements
Program Demand	• Number of applications requested per semester • Number of applications received per semester • Number of telephone contacts per semester
Student Satisfaction	• End of course student ratings of program content, quality of instruction, instructional resources, technology used, amount of interaction, instructional methods • Mid-semester student ratings • CISS data regarding perceptions of course interaction, course structure, and technical support
Faculty Satisfaction	• Faculty ratings of technology, technical support, interaction with students, quality of student work
Student Retention	• Percentage of dropouts from beginning to end
Student Learning	• Self-assessment of learning • Course project scores • Quiz and Test scores • Course grades
Financial Efficiency	• Course design cost • Course delivery cost • Publicity cost • Total unit cost, direct cost, overhead cost, • Tuition revenue

Activity 3: Knowledge Engineering

As work progressed in conceptualizing the evaluation system and identifying vital signs, alternatives for automating the system were explored. Literature on artificial intelligence, decision support software, electronic performance support systems (EPSS), and expert systems was reviewed by project staff. Due to the technology potential for the evaluation model, its focus on internal rather than external evaluation, and its purpose as a self-evaluation tool to be used by a variety of online personnel, including program administrators, instructors, and instructional designers, it was ultimately decided to utilize performance improvement technologies to automate much of the data collection and analysis functions. Performance improvement technologies make dynamic use of technology to facilitate data

collection and analysis tasks and to improve individual performance by providing timely information, advice, coaching, and training. An electronic performance support system is an example of a performance improvement technology that "captures, stores, and distributes individual and corporate knowledge assets throughout an organization to enable an individual to achieve a required level of performance in the fastest possible time and with the minimum of support from other people" (Raybould, (1995), p. 11). EPSS technologies often include an information database, an expert advisor, customized tools and templates, and the potential to run simulations. Beyond a few efforts in corporate settings, this technology has not been widely applied in education and training for evaluation purposes.

Available literature on performance improvement technologies was reviewed. A key component to any expert system is the knowledge base of a recognized expert in the problem area (Hayes-Roth, et al., 1983). Focus groups of recognized online education experts from the University of Illinois were held to ascertain their perceptions of what constitutes effective online education and evaluation. These experts worked through simulated evaluation problems to test their ideas and were asked to develop specific guidelines and suggestions for improving the vital sign assessment and in-depth procedures. The framework for the expert system was based on the outcomes of this activity.

Activity 4: EPSS Development

The investigative framework gleaned from the knowledge engineering activity was developed into a series of If-Then rule process charts (one for each vital sign) by the developers. One partial example of a vital sign rule process chart is shown in Table 2 where a variety of "then" statements are provided in response to the low indication of student satisfaction.

Table 1
Sample EPSS Rule Process Chart for One Vital Sign

Vital Signs	Data Elements
IF: Student Satisfaction as indicated by Student Rating Form is Low,	THEN: • review subscores for student ratings results. • identify the subscores that reveal the area of problem. • review results to individual items in the problem subscore(s).
IF: If more information is needed to determine the problem,	THEN: • develop an interview questionnaire and contact a sample of students by phone. • develop an e-mail questionnaire with relevant questions and send to a sample of students. • summarize new student data. • present the summary of all results to the instructor and discuss potential causes of the problem.

The probable causes identified by the online instruction stakeholders were incorporated into the rule process charts to complete the exploration diagrams. Sample computer display screens were also developed that simulate what each screen on the finished performance system program should look like. Using the rule process charts and sample screens, project staff created the performance support system. Prototype system routines were reviewed and revised to correct programming errors, to improve interface inadequacies, and to ensure user friendliness.

Activity 5: Field Testing

The performance support system shell was reviewed by selected professionals as a preliminary step to pilot testing. Data were programmed into the system to allow users to explore all possible investigative paths in the in-depth analysis. Formal and informal feedback was provided, which resulted in minor revisions in the programming (elimination of "bugs" from the system) and content of the evaluation system. The evaluation system is currently being tested within an online masters degree program in HRD at a large Midwestern university. Following this

initial testing, there are plans to expand the field testing to other university level online programs and in corporate settings.

Implications and Discussion

What truly separates the vital sign concept from traditional evaluation is its use of specific outcome measures with minimal data requirements to create a cursory picture of the general status of a program. The six vital signs are measured for each individual online course. Efficiency is maximized because initial data requirements are minimal, only specific programs are examined, and only problematic vital signs are subjected to further examination.

The most obvious technical feature of this new evaluation system is its use of the electronic performance support system. Used primarily in medicine, manufacturing, and engineering up to this point, this is the first attempt to adapt advanced technology to educational evaluation. This application, of course, has its limitations. Attempting to set limits on variables in social science research always causes consternation, for fear of eliminating some possible responses from consideration. In developing investigative paths for the expert system, limits had to be set. However, the data collection instruments and procedures were carefully developed and extensively field tested. The knowledge and skill of recognized evaluation and online experts were incorporated into the system. These factors, coupled with the extensive field testing still to be completed, result in an evaluation system that is as true to human expert evaluation as is possible. It is also important to remember that the system will be operated by human evaluators and all final judgments will be made by personnel of the program online. Thus, it should not be construed that the computer is replacing the human evaluator. It is only providing valuable assistance, which will facilitate self-evaluation of online programs.

The broad applicability of the vital signs gives this system utility for online program evaluation beyond the University of Illinois. Ultimately, the model developed for use in this system, along with the computer technologies that are applied, could be adapted for use in evaluating any online course or program in the public sector as well as in private sector training and human resource development programs.

References

Alkin, M. C. (1969). Evaluation theory development. *Evaluation Comment, 2*, 2-7.

Arenson, K. W. (1998, November 2). More colleges plunging into uncharted waters of online courses. *The New York Times*.

Clark, R. E. (1994). Assessment of distance learning technology. In E. L. Baker & H. F. O'Neil, Jr. (Eds.), *Technology assessment in education and training* (pp. 63-78). Hillsdale, NJ: Erlbaum.

Draper, S. W. (1996). Observing, measuring, or evaluating courseware [On-line]. Available: http://www.psy.gla.ac.uk/~steve/Eval.HE.html

Hayes-Roth, F., Waterman, D., & Lenat, D. (1983) *Building expert systems.* Reading, MA: Addison-Wesley.

Holton, E. F. III. (1995). In search of an integrated model for HRD evaluation. In E. F. III, Holton (Ed.), *Proceedings of the 1995 Academy of Human Resource Development Annual Conference* (pp. 4-2). Baton Rouge, LA: Academy of HRD.

Jackson, G. A. (1990). Evaluating learning technology: Methods, strategies, and examples in higher education. *Journal of Higher Education, 61*(3), 294-311.

Kaufman, R., & Keller, J. M. (1994). Levels of evaluation: Beyond Kirkpatrick. *Human Resource Development Quarterly, 5*, 371-380.

Khan, B. H. (1997). Factors to consider when evaluating a web-based instruction course: A survey. In B. H. Khan (Ed.), *Web-based instruction* (pp. 375-378). Englewood Cliffs, NJ: Educational Technology Publications.

Kimball, L. (1998). Easier evaluation with web-based tools. *Training & Development, 52*(4), 54-55.

Kirkpatrick, D. L. (1976). Evaluation of training. In R. L. Craig (Ed.), *Training and development handbook* (pp. 18-1-18-27). New York: McGraw-Hill.

Kirkpatrick, D. L. (1996). Evaluation. In R. L. Craig (Ed)., *Training and development handbook* (4th ed., pp. 294-312). New York: McGraw-Hill.

Kirkpatrick, D. L. (1998). *Evaluating training programs* (2nd ed.). San Francisco: Berrrett-Koehler.

Magalhaes, M. G. M., & Schiel, D. (1997). A method for evaluation of a course delivered via the world wide web in Brazil. *The American Journal of Distance Education, 11*(2), 64-71.

McCollum, K. (1998, May 15). Accreditors are urged to prepare to evaluate distance learning. *The Chronicle of Higher Education*, A34.

Mendels, P. (1999, January 6). Universities embrace technology, but distance learning faces controversy. *The New York Times*, Education Column.

Middleton, A. J. (1997). How effective is distance education?. *International Journal of Instructional Media, 24*(2), 133-137.

Nichols, G. W. (1997). Formative evaluation of web-based instruction. In B. H. Khan (Ed.), *Web-based instruction* (pp. 369-374). Englewood Cliffs, NJ: Educational Technology Publications.
Oakes, K. (1997). The hardest question to answer about CBT. *Training & Development, 51*(9), 45-47.
Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text* (3rd ed.). Thousand Oaks, CA: SAGE.
Pisik, G. B. (1997). Is this course instructionally sound? A guide to evaluating online training courses. *Educational Technology, 37*(4), 50-59.
Public Schools.
Raab, R., Swanson, B., Wentling, T. L., & Clark, C. (1992). *Training Evaluation*. Rome, Italy: Food and Agriculture Organization of the United Nations.
Ravitz, J. (1997). Evaluating learning networks: A special challenge for web-based instruction. In B. H. Khan (Ed.), *Web-based instruction* (pp. 361-368). Englewood Cliffs, NJ: Educational Technology Publications.
Raybould, B. (1995). Performance support engineering: An emerging development methodology for enabling organizational learning. *Performance Improvement Quarterly, 8*(1), 7-22.
Scriven, M. S. (1967). *The methodology of evaluation* (Perspectives of Curriculum Evaluation, and AERA monograph Series on Curriculum Evaluation, No. 1). Chicago: Rand NcNally.
Simonson, M. R. (1997). Evaluating teaching and learning at a distance. *New Directions for Teaching and Learning, 71*, 87-94.
Stake, R. E. (1967). The countenance of educational evaluation. *Teachers College Record, 68*, 523-540.
Stake, R. E. (1978). The case-study method in social inquiry. *Educational Researcher, 7*, 5-8.
Stake, R. E. (1981). Setting standards for educational evaluators. *Evaluation News, 2*(2), 148-152.
Stufflebeam, D. L. (1967). The use and abuse of evaluation in Title III. *Theory Into Practice, 6*, 126-133.
Stufflebeam, D. L., Foley, W. J., Gaphart, W. J., Guba, E. G., Hammond, R. L., Merriman, H. O., & Provus, M. M. (1971). *Educational evaluation and decision making.* Itasca, IL: Peacock.
Stufflebeam, D. L., & Shinkfield, A. J. (1985). *Systematic Evaluation.* Boston: Kluwer-Nijhoff.
Thorpe, M. (1993). *Evaluating open & distance learning* (2nd ed.). London: Longman.
Waterman, D. A. (1986). *A guide to expert systems.* Reading, MA: Addison-Wesley.
Wentling, T. L. (1980). *Evaluating occupational education and training programs.* Urbana, IL: Griffon Press.
Wentling, T. L. (1992). *Ameliorer La Qualite De La Formation*, Rome, Italy: Food and Agriculture Organization of the United Nations.
Wentling, T. L.(1993). *A trainer's guide to curriculum development*. Rome, Italy: Food and Agriculture Organization of the United Nations.