

EVALUATING VIRTUAL LEARNING

Romica TRANDAFIR(1)

Mihai Stefan NISTORESCU(2)

Ion MIERLUS-MAZILU(3)

Department of Mathematics and Computer Science

Technical University of Civil Engineering Bucharest

B-ul Lacul TEI, Nr. 122-124, Sector 2, Bucharest, ROMANIA

(1) romica@utcb.ro, (2) nist@utcb.ro, (3) mmi@utcb.ro

ABSTRACT

In considering literature on the evaluation of Virtual Learning's (VL) or similar technologies, it is apparent that there are many different approaches to studies. A useful framework has been devised by Oliver (1997), which provides a comprehensive guide to the evaluation of the use of educational technology.

It is possible that the term "evaluation" may be restrictive in the current context. Evaluation has been clearly explained by Oliver (2000) as "the process by which people make value judgements" and when applied to learning technology, he suggests that this is often the educational value of innovations or practical issues in introducing new teaching methods and resources. Whilst the overall objectives of such evaluations are likely to be identifying what may improve learning, some evaluations have specific outcomes, whilst others aim for more general relevance. Oliver (1997) is well aware of this distinction, which is built into the five purposes for evaluation (described in paper). A more marked distinction is made in the current paper by suggesting that it may be helpful to regard some studies as "experiments" and some as "evaluations".

1. INTRODUCTION (WHY EVALUATE?)

Institutional strategies for the development and use of ICT in Higher Education in the Romania is that new technologies should encourage rethinking of pedagogical aspects of teaching, learning and assessment. The Virtual Learning Environment Blackboard was purchased in 2002 and there are currently approximately 100 courses online.

Evaluation provides feedback for course developers on teaching and learning and is an important part of quality assurance. However, constraints on time and possibly expertise preclude most developers from detailed studies. Whilst it is still important to carry out evaluations of individual courses, looking for more general principles derived from experiments can provide guidance in the design and development of VLEs. Such research may also address issues that are not covered in many evaluations.

Effective teachers use a variety of means, some formal and others informal, to determine how much and how well their students are learning. For example, to formally evaluate student learning, most teachers use quizzes, tests, examinations, term papers, lab reports, and homework. These formal evaluation techniques help the instructor to evaluate student achievement and assign grades.

To evaluate classroom learning informally, teachers also use a variety of techniques. For example, teachers pose questions, listen carefully to student questions and comments, and monitor body language and facial expressions. Informal, often implicit evaluations permit the teacher to make adjustments in their teaching: to slow down or review material in response to questions, confusion, and misunderstandings; or to move on when student performance exceeds expectations.

When teaching at a distance, educators must address a different teaching challenge than when teaching in a traditional classroom. For example, instructors no longer have:

- A traditional, familiar classroom.
- A relatively homogeneous group of students.

- Face-to-face feedback during class (e.g. students' questions, comments, body language, and facial expressions).
- Total control over the distance delivery system.
- Convenient opportunities to talk to students individually.

For these reasons, distance educators may find it useful to not only formally evaluate students through testing and homework, but to use a more informal approach (see Angelo and Cross, 1993) in collecting data to determine:

- Student comfort with the method used to deliver the distant instruction.
- Appropriateness of assignments.
- Clarity of course content.
- If class time is well spent.
- Teaching effectiveness.
- How a course can be improved.

2. THE NATURE OF EVALUATIONS

In considering literature on the evaluation of VLs or similar technologies, it is apparent that there are many different approaches to studies. A useful framework has been devised by Oliver (1997), which provides a comprehensive guide to the evaluation of the use of educational technology.

It is possible that the term 'evaluation' may be restrictive in the current context. Evaluation has been clearly explained by Oliver (2000) as 'the process by which people make value judgements' and when applied to learning technology, he suggests that this is often the educational value of innovations or practical issues in introducing new teaching methods and resources. Whilst the overall objectives of such evaluations are likely to be identifying what may improve learning, some evaluations have specific outcomes, whilst others aim for more general relevance. Oliver (1997) is well aware of this distinction, which is built into the five purposes for evaluation (described below).

2.1 Purpose of evaluation

Roles

The starting point for distinguishing between different evaluations is naturally the purpose of the study. Oliver (1997), based on Draper, Brown, Henderson and McAteer (1996), identified five roles for evaluation: formative, summative, illuminative, integrative evaluations and quality assurance. Quality assurance is undoubtedly a specific purpose for evaluations. Explanations of illuminative and integrative evaluations illustrate the close relationship between purpose, approach (e.g. experimental versus ethnographic) and measures. For instance, illuminative evaluations are described as being primarily ethnographic, as opposed to experimental.

Experiments

Four of these five roles are identifying problems, describing and interpreting events, rather than studies, which may test a single well-defined question (summative evaluations) and provide results of more general relevance. These objectives provide criteria for distinguishing between evaluations and experiments. The specific research questions were whether the audio messages would increase the frequency of student participation and length of utterances in online asynchronous group discussion and whether they would also result in more favourable student perceptions.

These designs can be problematic in natural settings due to difficulties in achieving comparable situations, avoiding contact between groups where they may share material specifically intended for one group, and possible ethical problems such as depriving some people of a potentially richer learning environment.

Usability versus learning

Another dimension that separates studies is the approach adopted by the specific discipline. Whilst studies within the educational field aim to assess students learning outcomes, situating the evaluation within an educational context that incorporates assessment, an alternative objective is to measure usability of the system and its tools.

Definitions of usability vary but there are similarities in the type of variables they tend to measure. These include effectiveness, efficiency and satisfaction (ISO 9241), ease of remembering and error rate (Nielsen, 1993). Commonalities among definitions found in the literature are making the use of a system easier and more comfortable for the users, whilst guaranteeing a high level of productivity.

However to measure the level of productivity in the field of learning technologies may be particularly difficult. The crucial point is the conception of learning that underlies evaluation. Typical measures used to evaluate the usability of a system, response time, accomplishment of tasks, error rate, etc. are suitable for a large range of systems and even for Computer Assisted Instruction Systems (CAIS) or Intelligent Tutor Systems (ITS). However, if learning is conceived as a matter of process, during which a transformation of knowledge occurs, such measures say nothing about how new knowledge has developed and what is necessary to support this development.

As all activity within a VLE is carried out through the interface, it is important to examine how this may support learning. However, it is unhelpful to take the evaluation out of the learning context to focus only on ease of use of the system. The purpose of the evaluation should determine what is measured but it is the conception of the investigated phenomena that defines what is actually observed. In usability research the focus of the studies seems to be the individual using the system. Cultural factors that surround the use of the system are not included in the analysis. The context is merely a scenario that provides information about the task performed but is not part of the experience. Usability and learning may be combined in a single study, but each will have their own individual measures. How measurement is conducted is affected not only by the specific variables, but also by the circumstances surrounding the evaluation.

2.2. Types of Evaluation

Evaluation can be either formative, summative, or a combination of both.

Formative evaluation:

- Is an on-going process to be considered at all stages of instruction.
- Will enable the instructor to improve the course as he/she proceeds.
- Facilitates course and content adaptation.
- Will identify major gaps in the instructional plan or the need for minor adjustments.

Some strategies that educators can use to collect formative data from their distant students include:

- Post cards - provide each student with prestamped and preaddressed postcards. On a weekly basis, have students use the postcards to share their concerns or respond to questions during the last three to five minutes of class.
- Electronic mail - Can be a very effective way for instructors and students to communicate. Another plus, while the instructor is eliciting information about classroom learning, students become familiar with the use of electronic mail, a valuable skill.
- Telephone - Call students often. Ask them open ended questions (e.g., "What snags did you run into on the second writing assignment?") to let students voice their concerns. Follow with probes (e.g., "Then, will you need more information sources?"). Set phone-in office hours but be sure to welcome calls at other times.

Summative evaluation:

- Assesses overall effectiveness of the finished product or course.
- Can be a springboard in developing a revision plan.
- Can be a baseline of information for designing a new plan, program, or course.

- Will not help current students since it is conducted upon course completion.
- Some questions that educators may want to ask students when collecting summative data include:
- List five weaknesses of the course.
 - List three (or five) strengths of the course.
 - If you were teaching the course, what would you do differently?
 - Student background information: age, level in school, number of distance delivered courses taken prior to this one.
 - What would you recommend to a friend planning to take this course?
 - What did you think would be covered in this course but was not?
 - Would you recommend this course to a friend? Why or why not?

3. EVALUATION METHODS

Within the context of formative and summative evaluation, data may be collected through quantitative and qualitative methods.

Quantitative evaluation:

- Involves asking questions which can be statistically tabulated and analyzed, frequently using a scale, check list, or yes/no responses.
- Limits students to responding to the categories made available to them.
- Needs a large student sample for relevant statistical analyses.

Quantitative methods may be most useful for gathering information on large numbers of respondents for whom more in-depth, personalized approaches are not feasible. However, they do have some significant drawbacks:

- Many distance education courses have relatively small class sizes with students from various backgrounds. These small, stratified populations typically defy relevant statistical analysis.
- Quantitative surveys typically result in a rate of return of under 50 percent. A low rate of return often suggests that only those feeling very positively or negatively about the course responded to the evaluation.
- By definition and design, forced choice surveys offer respondents a limited number of possible response options. Therefore, fresh insights and unique perspectives falling outside the provided response categories go unreported.
- The cumbersome and often tedious nature of quantitative data collection can discourage formative evaluation, and often results in an over-reliance on summative evaluation.
- Statistical analysis often results in an illusion of precision that may be far from reality.

Qualitative evaluation:

- Is typically more subjective.
- Involves gathering a wider range and depth of information.
- Is more difficult to tabulate into neat categories.
- Will be less affected by typical small class size.
- Is a more flexible and dynamic method.
- Is not limited to pre-conceived topic of inquiry.
- Allows for student output of topics.

Can use:

- Open ended questioning -- with respondents asked to identify course strengths and weaknesses, suggest changes, explore attitudes towards distance delivery methods, etc..
- Participant observation -- with the distance educator observing group dynamics and behavior while participating in the class as an observer, asking occasional questions, and seeking insights regarding the process of distance education.

- Non-participant observation -- with the distance educator observing a course (e.g., an audioconference, interactive television class, etc..) without actually participating or asking questions.
- Content analysis -- with the evaluator using predetermined criteria to review course documents including the syllabus and instructional materials as well as student assignments and course-related planning documents.
- Interviews -- with a facilitator or specially trained individual collecting evaluative data through one-on-one and small-group interviews with students.

4. MEASURES AND WHAT TO EVALUATE

A sample of measures are briefly described to illustrate different approaches. In general, what is measured determines the type of data that needs to be collected, the stage of activity to focus on, and who provides the data. The measures are chosen to answer the research question (in the case of an experiment) or provide the appropriate feedback in an evaluation. Issues of usability can be addressed by looking at responses to the system and eliciting perceptions. Learning is generally assessed through outcomes, but perceptions may again be informative. There may also be interactions between the usability of the system and the nature and extent of learning. Therefore comparing participation in discussions may contribute to assessing the role of the interface in the facilitation of learning.

We can consider the following areas:

- Use of technology - familiarity, concerns, problems, positive aspects, attitude toward technology.
- Class formats - effectiveness of lecture, discussion, question and answer; quality of questions or problems raised in class; encouragement given students to express themselves.
- Class atmosphere - conduciveness to student learning.
- Quantity and quality of interaction with other students and with instructor.
- Course content - relevancy, adequate body of knowledge, organization.
- Assignments - usefulness, degree of difficulty and time required, timeliness of feedback, readability level of print materials.
- Tests - frequency, relevancy, sufficient review, difficulty, feedback.
- Support services - facilitator, technology, library services, instructor availability.
- Student achievement - adequacy, appropriateness, timeliness, student involvement.

Student attitude - attendance, assignments submitted, class participation.

Instructor - contribution as discussion leader, effectiveness, organization, preparation, enthusiasm, openness to student views.

Evaluation Tips

- Check out and adapt already published questionnaires; there's no need to re-invent the wheel.
- Draft and revise questions; change if necessary.
- Make use of follow-up probes:
- Alternate between instruction and interaction.
- Sequence your questions for best effect - go ahead and ask for suggestions for improvement before asking for what is good. This will help convey sincerity for seeking improvements.
- Place open ended questions after quick answer questions. This gives students built-in thinking time.
- On summative evaluation, assure anonymity. This can be accomplished by having all questionnaires sent to a neutral site where they would be removed from their envelopes and forwarded to the instructor without a postmark.
- Establish rapport by being interested and supportive. Withhold judgmental responses.
- Adapt to the student in degree of formality and pace of communication.<

- Use evaluation as a method for understanding teaching and learning.
- Try to get both positive and negative feedback. It is important not only to know what is not working, but also what is working.

5. CONCLUSIONS

Learner perceptions

A range of variables can be measured by asking learners for their perceptions. Attitudes are sometimes separated out from perceptions (e.g. Jones et al., 2000), but essentially both are measured by asking for an opinion or judgement. It is the focus of the question that differs. This may be satisfaction, estimates of how much they have learned, usefulness of tools in the VLE, etc.

Learning outcomes

These are an essential measure of a VLE that supports learning, but there can be difficulties in interpreting the results. As mentioned in 2.2.1, it may not be possible to attribute changes in outcomes to specific elements of a learning technology. Nevertheless, studies may provide indicators of variables which may be important and these can provide the basis for future experiments.

The particular aspect of performance that is measured is determined by the objectives of the course, and is therefore likely to vary across studies. However, if measurement is limited to the defined objectives, the evaluation may fail to identify other incidental learning which may take place. Oliver (1997) introduces a dimension labelled 'domain independence' which relates to this distinction. He points out that learning outcomes can be related to the specific subject, or be more generic, e.g. organising discussion. There may also be subject-specific outcomes which are not specified or anticipated by the teacher, but would be worth identifying.

120 REFERENCES

1. Angelo, T. & Cross, P. (1993). *Classroom assessment techniques: A handbook for college teachers*. San Francisco: Jossey-Bass Publishers.
2. Boyle, T and Cook, J. (2001) "Online interactivity: best practice based on two case studies", *Association for Learning Technology Journal*, Vol 9, No 1, pp 94–102.
3. Brown, S, Hardaker, CHM and Higgett, NP. (2000) "Designs on the web: a case study of online learning for design students", *Association for Learning Technology Journal*, Vol 8, No 1, pp 30–40.
4. Chang, C-C. (2001) "A study on the evaluation and effectiveness analysis of web-based learning portfolio", *British Journal of Educational Technology*, Vol 32, No 4, pp 435–458.
5. Dillon, A. (1992) "Reading from paper versus screens: a critical review of the empirical literature", *Ergonomics*, Vol 35, No 10, pp 1297–1326.
6. Nielsen, J. (1993) *Usability Engineering*, Academic Press, San Diego, USA.
7. Nielsen, J. (1994) "Enhancing the explanatory power of usability heuristics" in *CHI 94 proceedings*, B Adelson, S Dumais and J Olson (Eds), ACM Press, Reading, MA, pp 152–158.
8. Oliver, M. (1997) *A framework for evaluating the use of educational technology*, University of North London, Learning and Teaching Innovation and Development (LaTID) London.
9. Oliver, M. (2000) "An introduction to the evaluation of learning technology", *Educational Technology & Society*, Vol 3, No 4, pp 20–30.
10. Preece, J, Rogers, Y, Sharp, H, Benyon, D, Holland, S and Carey, T. (1994) *Human-computer interaction*, Addison Wesley, Wokingham, England, p603.
11. Richardson, J. (2001) "An evaluation of virtual learning environments and their learners: do individual differences effect perception of virtual learning environments", *Interactive Educational Multimedia*, Vol 3, No, pp 38–52.
12. Willis, B. (1993). *Distance education: A practical guide*. Englewood Cliffs, NJ: Educational Technology Publications.