

# EVALUATING THE EDUCATIONAL INFLUENCE OF AN E-LEARNING SYSTEM

Ani Grubišić, Slavomir Stankov, Branko Žitko

*Faculty of Natural Sciences, Mathematics and Education*

*Nikole Tesle 12, 21000 Split, Croatia*

*Phone: (385) 21-38 51 33-105, Fax: (385) 21-38 54 31*

*E-mail: ani.grubisic { slavomir.stankov, branko.bzitko}@pmfst.hr*

**Abstract:** Nowadays educational systems present their users (teachers and students) an intelligent environment in order to enhance the learning and teaching process. The goal of e-learning system developers is to build such systems that will create individualized instruction to get as close as possible to the 2-sigma boundary. Because of the fact that acquisition of knowledge is often an expensive and time-consuming process, it is important to know whether it actually improves student performance. In this paper we are going to present our approach about evaluating the educational influence of a e-learning system as well as some results on the evaluation of the e-learning system educational effectiveness in augmenting students' accomplishments for a particular knowledge domain by using the effect size as the metrics. By doing so, we determine whether and in which degree an e-learning system increases students' performance and can, therefore, be an adequate alternative for human tutors. *Copyright © 2002 IFAC*

**Keywords:** e-learning systems, evaluation, educational influence, effect size

## 1. INTRODUCTION

Evaluation is useful for investigation and exploration of the different and innovative ways in which technologies are being used to support learning and teaching. All instructional software should be evaluated before being used in educational process. Developers of the e-learning systems have become so involved in making their system work, that they have forgotten their original goal: to build an e-learning system that is as good or even better than highly successful human tutors. Moreover, they have paid little attention to the process of evaluation as they are required to be able to say something about the outcomes of an e-learning system. Since the major goal of an e-learning system is to teach, its

evaluation's main test is to determine whether students learn effectively from it (Mark and Greer, 1993). A useful definition of evaluation could be that evaluation is "providing information to make decisions about the product or process" (Phillips and Gilding, 2003).

A well-designed evaluation should provide the evidence, if a specific approach has been successful and of potential value to others (Dempster, 2004). It incorporates principles and methods used in other fields of educational or social science research.

Each methodology represents a different approach to evaluation. The fact that there are so many

Table 1 A brief history of e-learning systems evaluations (modified according to (Harvey, 1998))

<b>Decade</b>	<b>Evaluation</b>
1960s	Controlled, experimental studies. Learning is still regarded as independent of subject or context.
1970s	Still predominantly experimental process oriented descriptions. Methods include interviews, questionnaires, profiles, think aloud protocols, observations etc.
1980s	Experimental methods consistently fail to produce sufficient detail for designers - and evaluators' purposes in formative and summative studies. Usability studies take precedence over learning evaluation. Results of formative evaluation and various forms of user testing become important inputs to development, and the iterative design cycle is established.
1990s	Methods must accommodate situations where teachers and learners may never meet face to face. Evaluation is now accepted as an important and ongoing aspect of program and course improvement, the importance of context is undisputed. Part of an ongoing process which feeds back into a plan - implement - evaluate - improve loop. Studies involve qualitative and quantitative measures as appropriate.

approaches in common use simply reflects the fact that no single methodology is “the best”. Which one will be most appropriate for you depends on the type of questions you are asking. A unique model for evaluating e-learning systems is hard to define. Effective evaluation should include an examination of the pedagogical aspect and results of the learning and teaching process supported by evaluated e-learning system. It can help to ensure that learning technologies are developed and adopted in ways that support learners and teachers to realize their goals.

In this paper, we present a proposition of the e-learning systems evaluation methodology. We give an overview of existing evaluation methods as well as the methodology that can be used for evaluating the e-learning systems process.

## 2. EVALUATIONS METHODS AND INSTRUMENTS

Given the variety of educational system evaluation methods, it is not as easy to decide which one is appropriate in a particular context (Iqbal, *et al.*, 1999). Basically, there are two main types of evaluation methods (Frye, *et al.*, 1988): formative and summative. Formative evaluation focuses on improvements to products and processes which are being developed. It is often a part of a software engineering methodology where it is used to obtain information needed for modifying and improving a system’s functionality. The purpose of formative evaluation is to inform on-going processes and practices. It is important therefore that the findings are ready in time to enable you to make appropriate changes to your approaches or recommendations.

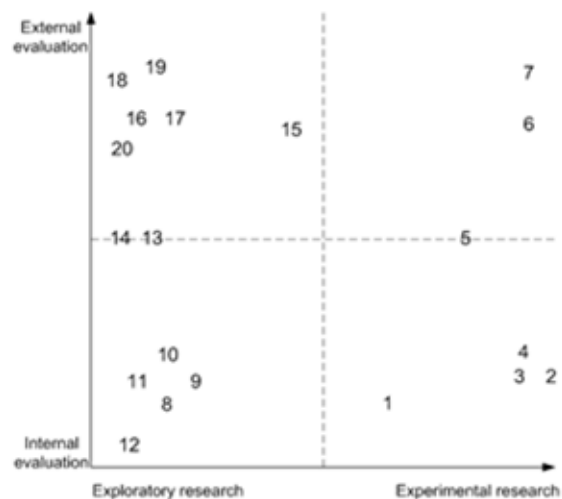
Formative evaluation doesn’t only concern itself with the e-learning system product, but also with the learning processes of students and our performance as teachers.

Summative evaluation is concerned with the evaluation of completed systems and tends to resolve, for e.g., such questions as: "What is the educational influence of an e-learning system on students?", "What does a particular e-learning system do?", "Does an e-learning system fulfill the purpose for which it was designed?", "Does an e-learning system result in predicted outcomes?" To summatively evaluate the effectiveness of e-learning system on student learning, we first need e-learning system which works in the way that it should. We also need to be clear about the type of learning the e-learning system is designed to achieve.

While planning an evaluation some of the tasks may need to be undertaken before you start development or implementation, such as collecting baseline information (pre-test data) for later comparison with currently existing conditions. Evaluation should be a planned, systematic but also open process; you should aim to incorporate opportunities for discovering the unexpected.

All evaluation methods, irrespective of their type, are classified along two dimensions (Fig. 1.) (Iqbal, *et al.*, 1999). The first dimension focuses on the degree of evaluation covered by the evaluating method. If the method only concentrates on testing a component of a system, it can be considered suitable for internal evaluation. If the method evaluates whole system, it is suitable for external evaluation. The second dimension differentiates between experimental research and exploratory research. Experimental research requires experiments that change the independent variable(s) while measuring the dependent variable(s) and require statistically significant groups. Exploratory research includes in-depth study of the system in a natural context using multiple sources of data, usually where the sample size is small and the area is poorly understood.

A well-designed evaluation incorporates a mix of techniques to build up a coherent picture. An evaluation answers the questions for which it was designed, hence the first step in research design is the identification of a research question. Hypotheses can be formed after identifying a research question, which must be testable, concerned with specific conditions and results, and possible to confirm or deny on the basis of those conditions and results. An evaluation methodology is then defined to enable the researcher to examine the hypothesis. When a practical, suitable evaluation method has been found



1. Proof of Correctness
2. Additive experimental design
3. Diagnostic accuracy
4. Feedback/instruction quality
5. Sensitivity Analysis
6. Experimental research
7. Product evaluation
8. Expert knowledge
9. Level of agreement
10. Wizard of Oz experiment
11. Performance metrics
12. Internal evaluation
13. Criterion-based
14. Pilot testing
15. Certification
16. Outside assessment
17. Existence proofs
18. Observation & qualitative classification
19. Structured tasks & quantitative classification
20. Comparison studies

Fig. 1. Classification chart for evaluation methods (Iqbal, *et al.*, 1999)

Table 2 Process of experimental research (modified according to (Harvey, 1998))

Phase	Description
Describe the intervention	Describe exactly what will be different in the students' experience after the change you propose as compared to the current situation.
Define the parameters	1. Only part of the class will experience the new learning situation, and their performance will be compared with that of their colleagues who have not experienced the change. 2. You plan to continue with your normal practice and compare the learning outcomes of your students with those who have experienced the new learning situation.
Define "success"	Decide what outcome would be needed for you to consider your experiment to be a success.
Decide how to measure successfulness	Decide how outcome can best be measured.
Analyze your data.	Analysis of data gathered through an experimental approach will most likely focus on deciding whether your innovation has had the predicted effect. Is there a difference to be seen in the outcome measure(s) gathered between your control and experimental situation? Is the difference in the direction which was predicted? And is the difference statistically significant? If it appears that differences do exist, then proceed to some test of statistical significance.

to answer the research question, the researcher can carry out the study and analyze data gathered through the study. Ideally, if results do not confirm the research hypothesis, researchers should be able to suggest possible explanations for their results.

The way in which you select your student sample will have an effect both on the information gathered and the impact that your findings might have. If you pick your own sample of students, you have the opportunity to select the students who are likely to be most co-operative or a group of students with the most appropriate skill levels. You can also select a random sample of students in order to try and get a more representative cross section from the class. You should watch that by selecting one group from a class and involving them in the evaluation study, you are not perceived as giving one group of students better support or tutoring than the rest of the class. It can happen that students complain about being discouraged from their peer group in some way (Harvey, 1998).

### 3. EVALUATING THE EDUCATIONAL INFLUENCE OF E-LEARNING SYSTEM

Experimental techniques are often used for summative research, where formal power is desired and where overall conclusions are desired. What is common in psychology and education (Mark and Greer, 1993), is that experimental research is suited to e-learning system because it enables researchers to examine relationships between teaching interferences and students' teaching results, and to obtain quantitative measures of the significance of such relationships.

Different evaluation methods are suitable for different purposes and the development of evaluation is a complex process. In a variety of different experimental designs, we have decided to

describe the usage of the pre-and-post test control group experimental designs that enable determining the effects of particular factors or aspects of the evaluated system.

Every educational innovation is an experiment in some sense of the word; you change something about the students' experience, predicting that better learning will take place. A controlled experiment is a way of teasing out the details of just which aspects of your innovation are influencing the outcomes you are considering and bringing about the changes you observe. The experimental method is a way of thinking about the evaluation process such that all the possible sources of influence are kept in mind.

#### 3.1 Pre and post testing

The idea of pre and post testing of students is often accepted as a viable instrument to assess the extent to which an educational intervention has had an impact on student "learning".

Pre and post testing is used because we know that students with different skills and backgrounds come to study a particular subject. We also need to establish a base measure of their knowledge and understanding of a topic in order to be able to quantify the extent of any changes in this knowledge or understanding by the end of a particular period of learning. Ideally, we wish to know not only that the educational intervention has had an impact on the student, hopefully a positive one, but we also want to be able to quantify that impact.

The process should require students who are undertaking a test to determine some individual starting level of knowledge or understanding of a topic. At a later point they should undertake the exactly comparable test to determine the extent to

Table 3 Process of pre and post testing (modified according to (Harvey, 1998))

Phase	Description
Test group	Student test group of at least 30 students.
Familiarization with e-learning system	Although an e-learning system might be simple to use it is important to ensure that students are familiar with all aspects of how to use the various features. You could consider organizing a familiarization session prior to your evaluation.
Pre and post testing	<ol style="list-style-type: none"> <li>work around the e-learning system Think about how much of the subject content they need to know before a pre test. Post test immediately after they have completed their study of the material in the e-learning system.</li> <li>selection of groups for two alternative modes of learning One group can use the e-learning system as a substitute for lectures (on at least 2 occasions). The second group can follow the standard lecture programme. Both groups should undertake pre and post tests.</li> <li>work around the lecture At this stage all students take the e-learning system unit prior to the delivery of the lecture in the topic. The pre and post testing is delivered immediately prior to and immediately after the lecture. These tests could be online or paper-based.</li> </ol>
Analysis of results	The various tests will provide a huge amount of data - some of it will be raw numeric data that can be analyzed using standard statistical tests.

which knowledge and understanding has been improved by the educational intervention.

The design of the pre and post questions is critical to success. The repetition of the same test questions is obviously not a sound solution to achieving comparability but it is a good idea to retain a proportion of the original test materials and to blend this with new questions which examine the same expected learning outcomes. It is also important to consider the type of questions which is used. Certainly we should not rely purely on objective questions. However, extended questions which seek to test a whole range of issues are also inappropriate.

### 3.2 Process of evaluation

For purposes of e-learning system evaluation, students that are picked to be part of experiment have to be randomly and equally divided into Control group and Experimental group. The Control group will be involved in traditional learning and teaching process and the Experimental group will use e-learning system. Both types of treatment should be scheduled for two hours weekly throughout one semester (2hr/week x 15 weeks = 30 hours/semester).

Both groups will take a 45-minute paper-and-pen pre-test that will be distributed at the very beginning of the course. Also, both groups will take a 60-minute paper-and-pen post-test that two weeks after the end of the course. Their results will be scored on a 0-100 scale. The pre-test enables to obtain information on the existence of statistically significant differences between the groups concerning student's foreknowledge. However, the post-test enables to obtain information on the existence of statistically significant difference between the groups concerning evaluation influence of the e-learning system.

### 3.3 Analysis of results

Data analysis techniques are best chosen in relation to the types of data you have collected. Quantitative data will rely on correlation and regression methods – 't' tests, analysis of variance, chi square as statistical outputs. Qualitative data may include transcripts from questionnaires, interviews or focus groups.

Interpreting the results of evaluation is difficult. In terms of the students' perception of the experience, for example, do students like it because it's new, or hate it because it's unfamiliar? You might ask would the student wish to use the e-learning system again and what improvements would they like to see. In terms of student performance, is it possible to isolate the effect of the new medium; is any change in scores the result of having a different group of student? Students will not always express their feelings, preferences, goals, or any changes in their study behaviors using the same words. There may be cultural or gender issues that influence what and how students say something. All these factors may distort the evaluation (Dempster, 2004).

The t-test is the most commonly used method to evaluate the differences between two groups. Since the primary intention of the e-learning system educational influence evaluation is to value the overall effectiveness and the effect size of e-learning system, so t-value of means of gains of test scores among the two groups have to be computed and compared (StatSoft, 2004).

The p-value reported with a t-test represents the probability of error involved in accepting our research hypothesis about the existence of a difference. The critical region is the region of the probability distribution which rejects the null hypothesis. Its limit, called the critical value, is defined by the specified significance level. The most

commonly used significance level is 0.05. The null hypothesis is rejected when either the t-value exceeds the critical value at the chosen significance level or the p-value is smaller than the chosen significance level. The null hypothesis is not rejected when either the t-value is less than the critical value at the chosen significance level or the p-value is greater than the chosen significance level. In the t-test analysis, comparisons of means and measures of variation in the two groups can be visualized in box-and-whisker plots. These graphs help in quickly evaluating and "intuitively visualizing" the strength of the relation between the grouping and the dependent variable.

First, it has to be checked whether groups' initial competencies were equivalent before comparing the gains of the groups. That means calculating the mean of pre-test score of both groups with their standard deviation. Then the t-values of pre-test means have to be computed to determine if there is reliable difference between two groups. Now, hypotheses have to be stated, for example: "There is a significant difference between the Control and the Experimental group".

Next, the gain scores from pre-test to post-test are to be compared. That means calculating the mean of both groups with their standard deviation. Then the t-values of means of gain scores have to be computed to determine if there is a reliable difference between the Control and the Experimental group. If there is statistically significant difference, it implies that e-learning system had a positive effect on the students' understanding of the domain knowledge. In other words, our hypothesis is accepted.

The effect size is a standard way to compare the results of two pedagogical experiments. Effect size can be calculated by using different formulas and approaches, and its values can diverge. In our approach to evaluating the educational influence of an e-learning system, the average effect size has to be computed in order to get a unique effect size that can be used in some meta-analysis studies. There are four types of effect size: standardized mean difference, correlation, explained variance, and interclass correlation coefficient, according to (Mohammad, 1998).

For determining group differences in experimental research, the use of standardized mean difference is recommended (Mohammad, 1998). The standardized mean difference is calculated by dividing the difference between experimental and control group means by the standard deviation of the control group. The following formula is used for the calculation of this standardized score:

$$\Delta = \frac{\overline{X_e} - \overline{X_c}}{s_c}, \quad (1)$$

where  $X_e$  = mean of the experimental group;  $X_c$  = mean of the control group;  $s_c$  = standard deviation of the control group. The mean or arithmetic average is the most widely used measure of central tendency, and the standard deviation is the most useful measure of variability, or spread of scores.

Effect sizes can also be computed as the difference between the control and experimental post-test mean scores divided by the average standard deviation. According to (Frye, et al., 1988) the effect size can be calculated using this formula:

$$\Delta = \Delta(\text{post-test}) - \Delta(\text{pre-test}). \quad (2)$$

Effect size can be calculated using different formulas and approaches, and its values can diverge. In our approach to evaluating the educational influence of a e-learning system, we propose computing the average effect size in order to get a unique effect size that can be used in some meta-analysis studies.

#### 4. CONCLUSION

As we have stated, all instructional software should be evaluated before being used in educational process. A unique model for evaluation of the e-learning systems is hard to define and methodology we have presented in this paper can ease the search.

Presented evaluation methodology for e-learning systems bases itself on experimental research with usage of pre-and-post test control group experimental designs. Pre and post testing is a practical instrument to appraise the amount of educational influence of a certain educational intervention.

When it comes to interpreting the results of evaluation, the t-test is the most commonly used method to evaluate the differences between two groups. First, it has to be checked whether groups' initial competencies were equivalent before comparing the gains of the groups. Next, the gain scores from pre-test to post-test are to be compared.

This evaluation methodology has been used to evaluate educational influence of the Web-based intelligent authoring shell Distributed Tutor Expert System (DTEEx-Sys) (Stankov, 2004). The DTEEx-Sys effect size of 0.82 is slightly less than 0.84, a standard value for the intelligent tutoring systems (according to (Fletcher, 2003)).

#### ACKNOWLEDGEMENTS

This work has been carried out within projects 0177110 Computational and didactical aspects of intelligent authoring tools in education and TP-02/0177-01 Web oriented intelligent hypermedial authoring shell, both funded by the Ministry of Science and Technology of the Republic of Croatia.

## REFERENCES

- Cook, J. (2002). Evaluating Learning Technology Resources, LTSN Generic Centre, University of Bristol
- Harvey, J. (ed.) (1998) Evaluation Cookbook. Learning Technology Dissemination Initiative, Institute for Computer Based Learning, Edinburgh: Heriot-Watt University.
- Dempster, J. (2004). Evaluating e-learning developments: An overview, available at: [www.warwick.ac.uk/go/cap/resources/eguides](http://www.warwick.ac.uk/go/cap/resources/eguides)
- Fletcher, J.D. (2003). Evidence for Learning From Technology-Assisted Instruction. In: *Technology applications in education: a learning view*, (H.F. O'Neal, R.S. Perez (Ed.)), Mahwah, NJ: Lawrence Erlbaum Associates, pp.79-99
- Frye, D., D.C. Littman and E. Soloway (1988). The next wave of problems in ITS: Confronting the "user issues" of interface design and system evaluation. In: *Intelligent tutoring systems: Lessons learned*. (J. Psotka, L.D. Massey, S.A. Mutter and J.S. Brown (Ed)), Hillsdale, NJ: Lawrence Erlbaum Associates
- Heffernan, N. T (2001) Intelligent Tutoring Systems have Forgotten the Tutor: Adding a Cognitive Model of Human Tutors, dissertation, Computer Science Department, School of Computer Science, Carnegie Mellon University.
- Iqbal, A., R. Oppermann, A. Patel and Kinshuk (1999). A Classification of Evaluation Methods for Intelligent Tutoring Systems. In: *Software Ergonomie '99 - Design von Informationswelten* (U. Arend, E. Eberleh and K. Pitschke. (Ed)), B. G. Teubner, Stuttgart, Leipzig, pp. 169-181.
- Patel, A. and Kinshuk (1996). Applied Artificial Intelligence for Teaching Numeric Topics in Engineering Disciplines, Lecture Notes in Computer Science, 1108, pp. 132-140.
- Mark, M.A. and J.E. Greer (1993). Evaluation methodologies for intelligent tutoring systems. *Journal of Artificial Intelligence and Education*, 4 (2/3), pp. 129-153.
- Mohammad, N.Y. (1998). Meta-analysis of the effectiveness of computer-assisted instruction in technical education and training, doctoral dissertation, Virginia Polytechnic Institute and State University, Blacksburg, Virginia
- Phillips, R. and T. Gilding (2003). Approaches to evaluating the effect of ICT on student learning. ALT Starter Guide 8, available at: <http://www.warwick.ac.uk/ETS/Resources/evaluation.htm>
- Stankov, S., V. Glavinić, A. Granić and M. Rosić () Intelligent tutoring systems—research, development and usage, *journal Edupoint-informacijske tehnologije u edukaciji*, 1/1
- Stankov, S., V. Glavinić, A. Grubišić (2004). What is our effect size: Evaluating the Educational Influence of a Web-Based Intelligent Authoring Shell?. In: *Proceedings INES 2004 / 8th International Conference on Intelligent Engineering Systems*, (S. Nedevschi, I.J. Rudas (Ed.)). Cluj-Napoca : Faculty of Automation and Computer Science, Technical University of Cluj-Napoca, 2004. 545-550
- StatSoft (2004). Inc. "Electronic Statistics Textbook", available at: <http://www.statsoft.com/textbook/stathome.html>

