Submitted to: The Encyclopedia of Information Science and Technology, vol. 4, pp. 2058-2062.
Idea Group, Inc.

Multimedia Evaluations based on Cognitive Science Findings

Eshaa M. Alkhalifa
University of Bahrain

INTRODUCTION

Multi-Media systems waltzed into the lives of students and educators without allowing for the time required for the development of suitable evaluation techniques. Although everyone in the field is aware that judging this type of teaching software can only come through evaluations, the work done in this regard is scarce and ill-organized. Unfortunately, in many of the cases the evaluation forms were just filled in by instructors who pretended they are students when they went through the tutorial systems (Reiser et al., 1994).

BACKGROUND

In the early days, some researchers regarded the evaluation of the program's functional abilities and efficiency to be important so they defined them as formative evaluation and they also defined the effectiveness of the system as summative evaluation. (Scriven,1967; Bloom, et al.,1971).

Others believe the evaluation of the system is unimportant so they focused on the latter by comparing student performance in pre and post test questionnaires prior and following the use of the system, learning style questionnaires that targeted their learning preferences and a subjective questionnaire that investigated if they like the system (Kinshuk et al., 2000). Unfortunately, many of the pre and post tests resulted in no significant[1] differences in student grades when multi-media is compared to classroom lectures or to carefully organized well illustrated text books (Pane, et al., 1996). These disappointing results caused researchers to question whether or not the correct evaluation questions are being asked; for example should the test be of interactivity versus lack of interactivity, or should one compare animation with textual media (McKenna, 1995). If Pane et al., (1996) were aware of the work done by Freyd (1987) who studied the cognitive effects of exposing subjects to a series of still images to find that they are equivalent in the reactions they elicit to being exposed to a moving picture, then perhaps they would

[1] This term is defined in the Analysis of Results section.

not have asked whether animation is equivalent to a textbook with carefully set images of all stages.

Since the problem that arose is the evaluation question, researchers continued to alter it in order to recognize what should be emphasized. Tam et. al., (1997) proposed a three part evaluation procedure that includes peer review, student evaluation as well as pre- and post-testing (Tam, et. al., 1997). They were not able to get rid of the pre and post test evaluation as it is the primary test for how much learning was achieved and they still got no significant differences.

At this stage, researcher recognized that evaluations did not target the appropriate level of detail so Song et al. (2000; 2001) presented empirical support that animation helps reduce the cognitive load on the learner. They also showed that multi-media is more effective in teaching processes than in teaching conceptual definitions while textual presentations are better at the latter. However, all this was done in very limited test domains that lacked the realistic world of an educational system. Albalooshi and Alkhalifa (2002) implemented some of these ideas in addition to offering both textual representations and animations within the same screen to students. This supports individual learning preferences while offering multi-media systems as a cognitive tool. Such a tool requires an evaluation framework that is well informed of the justification behind its design and the way its main modules interact.

A 3-DIMENTIONAL FRAMEWORK FOR EVALUATION

In the reported cases, most of the evaluated systems failed to reflect their true abilities because some aspects of the design or effects were neglected. Consequently, a complete framework of evaluation is required to take into account all issues concerning the software and the learning process. Evaluation questions can be channeled into three main dimensions of evaluation that could then be subdivided into the various methods that form possible criteria that guides the evaluation process.

Table 1: A Three Dimensional Framework of Evaluation

1st Dimension: System Architecture

This dimension is concerned with the system's main modules, their programming complexity as well as their interactions. Evaluation within this dimension should be performed in any or all of the following methods:

- Full description of system modules and complete check of interaction.
- Expert survey of the system filled by experts or educators.
- Student evaluations to consider their perspective of the system.
- Architectural design must be based on cognitive science findings rather than chance.
- Everything else concerning the system design such as cost analysis and portability.

2nd Dimension: Educational Impact

This dimension is concerned with assessing the benefits that could be gained by students when they use the system. Classically, these are done in pre and post tests and this is carried on in this framework with more attention given to detail.

- Students grouped according to their mean grade in a quiz.
- Post tests are used to compare one group with system only and another classroom only. A third group attends the classroom lecture with the class group and does a pre-test then uses the system before doing a post test for comparison with the other two.
- Questions in the pre/post tests must be mapped to each other to test the same types of knowledge, mainly consisting of declarative and procedural knowledge.
- The tests should best be attempted with students who were never exposed to this material previously to assess their learning rate.

3rd Dimension: Affective Measures

This dimension is mainly concerned with student opinions on the user friendliness of the system and allows them to express any shortcomings in the system. This could best be done through a survey where students are allowed to add any comments they wish freely and without restraints.

The framework will be explained through a case study that was performed of a Data Structure Tutorial system (DAST) that was developed and evaluated at the University of Bahrain (AlBalooshi and Alkhalifa, 2003). The process started with a pre-evaluation stage where students were all given a test and then were divided into groups of equivalent mean grades. This was done to allow each group to have members of all learning levels.

Then the pre and post tests were written to ensure that one set of questions mapped onto the next by altering their order while ensuring they include declarative questions that require verbalization of how students understand concepts as well as procedural questions that test if students understand how the concepts can be applied. Last but not least, a questionnaire was prepared to allow students to highlight what they regard as any weak areas or strong areas based upon their interaction with the system. The evaluation procedure for students is shown in Figure 1. Educators were also asked to fill in an evaluation form as experts.

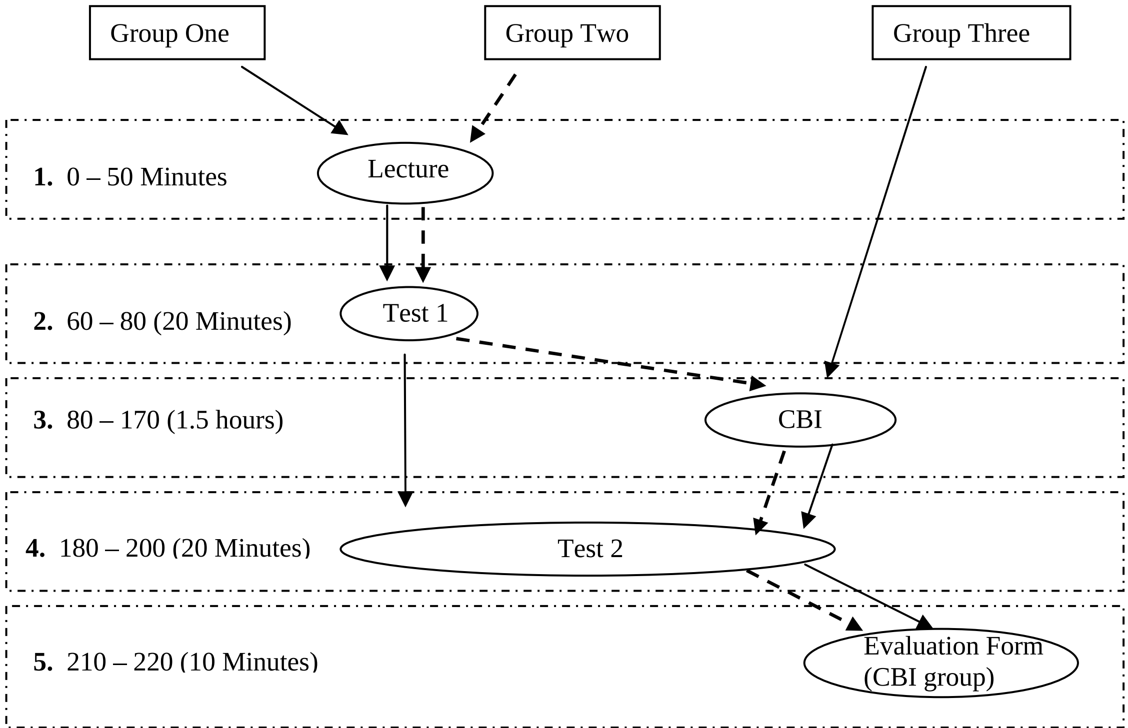

Figure 1. The Evaluation Procedure.

ANALYSIS OF RESULTS

First of all, student grades were analyzed using the Analysis of Variance (ANOVA) test. This test allows the evaluation of the difference between the means by placing all the data into one number, which is F, and returning as a result one p for the null hypothesis. It will also compare the variability that is observed between conditions, to the variability observed within each condition.

The statistic F is obtained as a ratio of two estimates of students variances. If the ratio is sufficiently larger than 1, then the observed differences among the obtained means are described as being statistically significant. The term "null hypothesis" represents an investigation done between samples of the groups with the assumption that additional learning will not occur as a result of the treatment. In order to conduct a significance test, it is necessary to know the sampling distribution of F given that the significance level needed to investigate the null hypothesis. It must be also mentioned that the range of variation of averages is given by the standard deviation of the estimated means.

The ANOVA test, did indeed show that there is a significant improvement in Group Two between the first test which was taken after the lecture and the second test which was taken after using the system. However, this was not sufficient to be able to pinpoint the strengths of the system. Therefore, a more detailed analysis was done of student performance in the individual questions of test one and test two. Since the questions were mapped onto each other by design, it was easy to identify significant changes in student grades in a particular question type for students of group two who responded to similar questions before and after the use of the system. For example, a highly significant improvement with F=58 and p<.000 was observed in the question "*Using an example, explain the stack concept and its possible use?*" which is an indication that the use of the system did strongly impact the student understanding of the concept of a "stack" in a functional manner.

Another point of view is to examine the scores by using the total average, which is 10.639, which can be approximated to 10.5 which can be used as a border-line. The rest of the scores can then be divided around this line. It was noticed that the average score of the third group was not high, yet 10 of scores were above the border-line while comparatively 6 scores were above it from the second group and only 6 of group one which took the class only option. This shows the results of the third group who took the CBI package alone and the second group which had both the classroom lecture and the CBI package exposure to be close. It also underlines how much the second group improved their test results after taking the CBI and in the same time showing that the first group had not improved much only with the lecture learning. Alkhalifa (2001) tested in abstract logical context the effects of describing a state of a "*moving system*" versus describing it as a static system to find that this affects logical conclusions subjects eventually arrive at. This tutorial, implemented those findings through this particular statistic revealed the effects of having animation in this multi-media system would.

These results indicate that the use of the system may introduce a "limiting" effect that follows the initial introduction to the concepts (Albalooshi and Alkhalifa, 2002). Classroom lectures introduce students to the concepts allowing them all the freedom to select all types of applications, which is in some ways overwhelming. The use of the system, on the other hand, produces a safe haven to test their ideas and strongly pursue the examples they can imagine which helps them arrive at a solid procedural understanding of the concepts. It goes without saying that such a conclusion would have been impossible to make if the questions were not purposely set in the shown mapped fashion.

Additionally, students of groups two and three who were exposed to the system, were asked to fill in an evaluation form composed of a series of questions as proposed by Caffarella (1987). They generally gave ratings, of around 4 to 5 on a scale that went 0 to 6 with the highest for "The use of graphics, sound, and color contributes to the student's achievement of the objectives" "The user can control the sequence of topics within the CBI program." And the lowest score which was 3.559 for "The level of difficult is appropriate for you". Therefore, it seems that the students in general enjoyed learning through the system although they found the level of difficulty of the concepts presented challenging.

In addition, to all this, three peer experts filled in evaluations forms to rate the system from an instructor's point of view and they gave the system an average rating of 5.33 on the same scale of 0 to 6.

FINE GRAINED EVALUATION

The evaluation framework proposed here evaluates multi-media educational software at a finer level of detail than what was previously followed. The system architecture for example, is evaluated in terms of the cognitive assumptions on which it relies. In many cases this dimension was over looked as in the studies that found animation and carefully organized images to be equivalent (Lawrence, et al., 1994; Byrne et al., 1999) because they were not informed of the work done to show that a sequence of images are translated as animation (Freyd, 1987). In fact, there is a difference between the two in that animation presents subjects with a cognitive tool that reduces the cognitive load they endure while learning and allows them to concentrate on the transfer of information rather than assimilating the presented material into animation (Albalooshi and Alkhalifa, 2002).

The educational impact question must also be carried out at a finer level of detail than just to report total grades for comparison. Specific questions may be designed to probe key types of learning as was shown in the presented case study.

FUTURE TRENDS

Only through a fine grained analysis can the true features of the system be revealed. It is this type of information that allows designers to take advantage of the opportunities offered by an educational medium that offers to transform the learning experience into a joy for both students and educators alike. Consequently, cognitive science concepts represent themselves to future research as a viable means to comprehend the learning process. This would guide the evaluative process to focus on the points of strengths and weaknesses of each system rather than assess it as a unit whole.

CONCLUSION

A three dimensional framework is presented as a means to evaluating multimedia educational software in order to resolve the short-comings the current evaluation techniques. It differs from the other, in that it seeks a more fine grained analysis while being informed through cognitive science findings. The extra detailed review reveals specific strengths of the system that may have been otherwise concealed if the comparison was done in the traditional manner. In other words, this process focuses on students' cognitive interpretations of what they are presented with rather than assuming to know how they will approach the educational content presented to them.

REFERENCES

Al Balooshi, F. & Alkhalifa, E. M. (2002), Multi-Modality as a Cognitive Tool, *Special Issues: Innovations in Learning Technology*, T. Okamoto & R. Hartley

(Eds), Journal of International Forum of Educational Technology and Society, IEEE, v. 5(4), pp.49-55

Alkhalifa, E. M. & Al Balooshi, F.(2003), A 3-Dimensional Framework for Evaluating Multimedia Educational Software, *Virtual Education: Cases in Learning & Teaching Technologies*, Fawzi Al Balooshi (Ed), Idea Group Publishing. (invited)

Alkhalifa, E. M.(2001), Directional Thought Effect in the Selection task, *Proceedings of the Third International Conference on Cognitive Science, ICCS* 2001, pp. 171-176, Beijing, China

Bloom, B.S. Hastings, J.T. & Madaus, G.F. *Handbook on formative and summative evaluation of learning*. New York: McGraw-Hill, 1971.

Byrne, M. D., Catrambone, R., Stasko, J. T. (1999), Evaluating animation as student aids in learning computer algorithms, *Computers and Education*, v. 33(4) pp. 253-278.

Caffarella, E. P. (1987). Evaluating the New Generation of Computer-Based Instructional Software. Educational Technology, 27(4), 19-24.

Freyd, J. (1987). Dynamic Mental Representations. *Psychological Review*, v.94(4), pp.427-438.

Kinshuk, Patel, A. & Russell, D. (2000), A multi-institutional evaluation of Intelligent Tutoring Tools in Numeric Disciplines, Educational Technology & Society Volume 3, Part 4, http://ifets.ieee.org/periodical/vol_4_2000/kinshuk.html.

Lawrence, W., Badre, A. N. and Stasko, J. T. (1994), Empirically Evaluating the Use of Animation to Teach Algorithms, Georgia Institute of Technology, Technical Report GIT-GVU-94-07, Atlanta, 1994.

McKenna, S. (1995) “Evaluating IMM: issues for researchers.” In *Occasional Papers in Open and Distance Learning*, No 17. Open Learning Institute, Charles Sturt University.

Pane, J. F., Corbett, A. T., & John B. E. (1996), Assessing Dynamics in Computer-Based Instruction, *In Proceedings of the 1996 ACM SIGCHI Conference on Human Factors in Computing Systems*. Vacouver, B.C. Canada, April 1996, pp.197-204.

Reiser, R.A., & Kegelmann H. W. (1994). Evaluating Instructional Software: A Review and Critique of Current Methods. *Educational Technology, Research and Development*. Volume 42, Part 3, pages 63-69, USA.

Scriven, M. The methodology of evaluation. In R.E. Stake (Ed.) *Curriculum evaluation*, Chicago: Rand-McNally, 1967.

Song, S. J., Cho, K. J. & Han, K. H. (2000). The effectiveness of cognitive load in multimedia learning. The Proceedings of the Korean Society for Cognitive Science, pp. 93-98.

Song, S. J., Cho, K. J. & Han, K. H. (2001). Effects of presentation condition and content type on multimedia learning, Proceedings of the Third International Conference on Cognitive Science, ICCS 2001. Beijing, China. pp. 654-657.

Tam, M., Wedd, S., & McKerchar, M. (1997). Development and Evaluation of a Computer-Based Learning Pilot Project for Teaching of Holistic Accounting Concepts. *Australian Journal of Educational Technology*. Volume 13, part 1, pages 54-67, Australia.

Terms and Definitions

Cognition: The psychological result of perception, learning and reasoning.

Cognitive Load: The degree of cognitive processes required to accomplish a specific task.

Cognitive Science: The field of science concerned with cognition and includes parts of cognitive psychology, linguistics, computer science cognitive neuroscience and philosophy of mind.

Cognitive Tool: A tool that reduces the cognitive load required by a specific task.

Declarative versus Procedural Knowledge: The verbalized form of knowledge versus the implemented form of knowledge.

Learning Style: This is the manner in which an individual acquires information.

Multimedia System: Any computer delivered electronic system that presents information through different media that may include text, sound, video computer graphics and animation.