Multimedia Evaluations Based on Cognitive Science Findings

Eshaa M. Alkhalifa

University of Bahrain, 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 were 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 (Bloom et al., 1971; Scriven, 1967).

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 to and following the use of the system, learning style questionnaires that targeted their learning preferences and a subjective questionnaire that investigated whether students like the system (Kinshuk et al., 2000). Unfortunately, many of the pre- and post-tests resulted in no significant1 differences in student grades when multi-media is compared to classroom lectures or to carefully organized, well-illustrated textbooks (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 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, researchers 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-DIMENSIONAL 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 guide the evaluation process.
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 & 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.

**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 hypoth-
Related Content

Humanware Issues in a Government Management Information Systems Implementation
[www.igi-global.com/chapter/humanware-issues-government-management-information/44537?camid=4v1a](www.igi-global.com/chapter/humanware-issues-government-management-information/44537?camid=4v1a)

Adoption of E-Commerce in the Value Chain by SMEs
[www.igi-global.com/chapter/adoption-commerce-value-chain-smes/14212?camid=4v1a](www.igi-global.com/chapter/adoption-commerce-value-chain-smes/14212?camid=4v1a)

A Framework for Improving Effectiveness of MIS Steering Committees
[www.igi-global.com/article/framework-improving-effectiveness-mis-steering/50965?camid=4v1a](www.igi-global.com/article/framework-improving-effectiveness-mis-steering/50965?camid=4v1a)

Experiences in Social Innovation: A Platform for Ethics Through a School of Engineering Studies
Domingo Alfonso Martín Sánchez and Ana García Laso (2014). *Teaching Cases Collection* (pp. 4-17).
[www.igi-global.com/article/experiences-in-social-innovation/115955?camid=4v1a](www.igi-global.com/article/experiences-in-social-innovation/115955?camid=4v1a)