

A Cognitively-Based Framework for Evaluating Multimedia Systems

Eshaa M. Alkhalifa

University of Bahrain, Bahrain

INTRODUCTION

Multimedia systems waltzed into the lives of students and educators without allowing anyone 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 to be students when they went through the tutorial systems (Reiser & Kegelmann, 1994). Nowadays, however, awareness of the impact of evaluation results on the credibility of the claims made is rising.

BACKGROUND

Ever since the early days, researchers recognized the existence of two main dimensions of multimedia evaluations. Formative evaluation is concerned with the program's functional abilities and efficiency, while summative evaluations are concerned with the effectiveness of the system in achieving its goals (Bloom, Hastings, & Madaus, 1971; Scriven, 1967).

Heller and Martin (1999) inform us that the evaluation question depends on the core subject from which it emerges, and they present a list of four subjects, namely, computer science, computer graphics, education, and human-computer interaction. They explain that if the core subject is computer science, then the research question concerns the technical requirements of the multimedia systems, including data compression, storage requirements, bandwidth, and data transmission. If the question is from computer graphics, then the focus is on speed of image rendering, representation of light, and creation of animation. If the question is from education, then media is evaluated in terms of its impact on teaching and learning along with attributes such as motivation, feedback, and information delivery. If the question is from human-computer interaction, then it is concerned with the use of multimedia in interface design focusing on issues that impact the interaction itself, such as screen design, the use of metaphor, and navigational strategies.

In retrospect, when examining their classification, we find that the main two dimensions are well covered. Formative evaluation is the focus of computer science and computer

graphics as a whole, while summative evaluation is the main, but not only, focus of education and human-computer interaction. With education, we find that "motivation" does not conform to any of the two dimensions, while human-computer interaction requires both formative evaluation in addition to summative evaluation.

Researchers tested their systems through a summative evaluation frequently using a pretest and posttest where the first is taken prior to using the system and the second following the use of the system. Unfortunately, this type of testing has been plagued with no significant¹ differences in student grades when multimedia is compared to classroom lectures or to carefully organized, well-illustrated textbooks (Pane, Corbett, & John, 1996). Others widened the scope of their evaluation procedure by adding learning style questionnaires that targeted student-learning preferences and a subjective questionnaire that investigated motivation issues (Kinshuk, Patel, & Russell, 2000).

Disappointment in the results of pretests and posttests caused researchers to alter the main summation evaluation question. They wondered if the test is for the educational effects 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.

The problem that emerged in the summation dimension is therefore the question itself. Tam, Wedd, and McKerchar (1997) proposed a three-part evaluation procedure that includes peer review, student evaluation, and pre- and post-testing. They were not able to get rid of the pretest and posttest evaluation, as it is the primary test for how much learning was achieved, and they still saw no significant differences in their results. In other words, they collected more subjective feedback from users, and this is not classified under summation evaluation.

At this stage, researchers recognized that evaluations did not target the appropriate level of detail that would enable them to detect differences that may exist in their results. Song, Cho, and Han (2000, 2001) presented empirical support that animation helps reduce the cognitive load on the

learner. They also showed that multimedia 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. Al Balooshi and Alkhalifa (2002) presented such an educational system evaluation by showing that the student cognitive styles do affect how they learn and that only the correct research question is capable of detecting these differences.

The evaluation framework therefore has to be expanded to include the motivation measure that is tested by designers but does not conform to any predefined dimension and to include these more detailed research questions at the appropriate level of abstraction to detect the differences between the multimedia systems subject to evaluation.

A THREE-DIMENSIONAL FRAMEWORK FOR EVALUATION

All evaluations start with an evaluation question that compares the one thing to another. If the words used to classify the question are not distinctive enough, then research will not benefit from the findings or misclassify them. Motivating students, for example, does not necessarily imply that the educational impact will be influenced. Asking a question that

is at a high level of abstraction, like if one media teaches as well as another without specifying the details of the design, difference, and materials, may also generalize to the degree that true benefits are shadowed or lost in the generalization. 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.²

The framework is composed of three main steps that will be explained through a case study of a Data Structure Tutorial System (DAST) that was developed and evaluated at the University of Bahrain (Alkhalifa & Al Balooshi, 2003).

The first dimension was first reviewed by distributing three instructor surveys based on a series of questions proposed by Caffarella (1987), to allow them to illuminate the various anomalies that may be present in the multimedia system design. A similar evaluation was distributed among students to allow them to evaluate the system subjectively.

The second dimension required a slightly more complicated process, which started with a pre-evaluation test of all students to ensure they were divided into groups of equivalent mean grades. Then the pretests and posttests were written to ensure that one set of questions mapped onto the next by altering their order while ensuring they included declarative questions that required verbalization

Table 1. A three-dimensional framework for evaluation

<p>First Dimension: System Architecture</p> <p>This dimension is concerned with the system's main modules, their programming complexity, and their interactions. Evaluation within this dimension should be performed in any or all of the following methods:</p> <ul style="list-style-type: none">• Operation of the system as a whole is described and evaluated to show optimization techniques used, and so forth.• Expert survey of the system filled by experts in the field or educators.• User evaluations of the way the system works, to indicate if they ran into any errors or problems that have not been predicted by designers.• Architectural design of the system is presented and evaluated against prior work and evaluations of similar systems to ensure that it benefits from the lessons learned.• All business related issues such as cost analysis and portability, time frame required for mass production of similar systems, and so forth. <p>Second Dimension: Cognitive Impact</p> <p>This dimension is concerned with assessing the benefits that could be gained by students when they use the system. Classically all the following methods must be measured using pretests and posttests of educational impact prior to and following use of the system. Here we find four areas of focus: two types of knowledge, cognitive traits, and the classification of the materials presented.</p> <ul style="list-style-type: none">• Tests of declarative knowledge that required verbalization of what users understood or learned following the use of the system.• Tests of procedural knowledge that tested if students understood how the concepts could be applied in novel situations.• Tests of cognitive styles that impacted how well students learned from one approach to presenting information vs. an alternative approach.• Tests of the alignment of the materials taught with the teaching style selected for that material. For example, is teaching mathematics more effective if it is interactive with live computation or if it is through textual presentation. <p>Third Dimension: Affective Measures</p> <p>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 surveys where students are allowed to add any comments they wish freely and without restraints.</p>

3 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/cognitively-based-framework-evaluating-multimedia/13632

Related Content

The Expert's Opinion

Karen D. Walker (1995). *Information Resources Management Journal* (pp. 35-36).

www.irma-international.org/article/expert-opinion/51005

The eAssessment of Key Competences and their Relationship with Academic Performance

Fernando Martinez-Abad, Patricia Torrijos-Fincias and María José Rodríguez-Conde (2016). *Journal of Information Technology Research* (pp. 16-27).

www.irma-international.org/article/the-eassessment-of-key-competences-and-their-relationship-with-academic-performance/172089

Information Systems Research Relevance

Shirish C. Srivastava and Thompson S.H. Teo (2009). *Encyclopedia of Information Science and Technology, Second Edition* (pp. 2004-2009).

www.irma-international.org/chapter/information-systems-research-relevance/13853

ICT and Developing Social Capital

Kate Sutcliffe (2008). *Information Communication Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 672-678).

www.irma-international.org/chapter/ict-developing-social-capital/22694

Development of KABISA: A Computer-Based Training Program for Clinical Diagnosis in Developing Countries

Jef Van den Ende, Stefano Laganà, Koenraad Blot, Zeno Bisoffi, Erwin Van den Enden and Louis Vermeulen (2005). *Journal of Cases on Information Technology* (pp. 136-145).

www.irma-international.org/article/development-kabisa-computer-based-training/3143