Chapter 8
Evaluation of Interactive Multimedia Packages

INTRODUCTION

The literature shows that many different evaluation methodologies for computer aided learning (CAL) packages have been proposed based on different philosophical views (Worthen and Sanders, 1973; Popham, 1974; Stephen and Stanley, 1985). The evaluation may be used for a variety of purposes such as refining goals, defining products or programs, and estimating costs, usability and effectiveness (Reeves, 1993). This involves the systematic review of the content, design, and instructional value and worth of computer aided learning packages. In general, any instructional software package should be evaluated before it is delivered or used in the classroom or research laboratory. This Chapter provides some general evaluation techniques used in the evaluation of such packages.

EVALUATION TECHNIQUES

The lack of controlled evaluation procedures for computer-based instruction has led to disagreement regarding their success in student learning (Baker, 1990). There is an enormous uncertainty in the edu-

DOI: 10.4018/978-1-60566-764-5.ch008

cation society regarding the effectiveness of any evaluation scheme for software (Borich and Jernelka, 1981; Bates, 1981; Tucker 1989; Micceri *et al*., 1989). These educational technology practitioners suggest that evaluation schemes of CAL packages need further research, improvement and are highly subjective since they depend on the objectives of the designer and the context of their use.

Evaluation has generally been conceptualized as either formative or summative. The aim of the formative evaluation is the refinement and improvement of a program or learning package while the aim of the summative evaluation is to determine the impact and the outcomes of a particular program or learning package (Guba and Lincoln, 1991). A program or learning package that is acceptable for some initial evaluation should be tested for its efficiency in real environments. Reja (2003), in an effort to identify a standard evaluation procedure for CAL packages in Europe, conducted a study on 19 organizations within ten member states of the European Community and found diversity in both formative and summative evaluation activities. According to Reja's findings, formative evaluations, conducted by most organizations, identify weaknesses in a product early enough to implement design changes. The summative evaluations, which are conducted at the end of a major development to assess the various aspects of a finished product, varied from a critical appraisal of a product by an expert to extensive in-depth testing.

Shute and Regian (1993) provide a framework within which an experimental evaluation methodology for intelligent tutoring systems (ITS) could be standardized. While ITS involve computer-based tutoring that is based on artificial intelligence models, the proposed framework is also appropriate to the evaluation of interactive CAL packages (Muramatsu *et al.,* 1998). Baker and King (1993) proposed an evaluation methodology based on a checklist in which they identified "hallmarks of quality" that characterize good learning products. The authors defined twelve basic categories that embody good learning design: engagement, interactivity, tailor ability, appropriateness of multimedia mix, mode and style of interaction, quality of interaction, quality of end user interface, learning styles, monitoring and assessment techniques, built-in intelligence, adequacy of ancillary learning support tools, and suitability for single user/group/distributed use. This evaluation checklist was used to assess 43 wide ranging of computer-based learning and training products. Baker and King (1993) found that the quality of end-user design interface was very important in producing a quality product. Engagement, interactivity, as well as tailor ability were found to be the other benchmark of quality.

However, Eibeck (1996) pointed out that the shortcoming of the checklist suggested by Baker and King (1993) is that it only addresses student learning on a superficial level. Research in the area of cognitive science has indicated that certain learning models present within CAL packages have a greater potential to improve learning than others.

A number of methods could be used to determine the efficacy of a program or learning package. These include quantitative and qualitative methods and laboratory testing method. According to Legree *et al.* (1993) quantitative methods are preferred to determine overall effectiveness of a program or learning package. On the other hand Murray (1993) recommended that the qualitative methods are suitable to find the internal efficiency of the overall program or learning package and its individual components. However Wyatt and Spiegelhalter (1990) suggested laboratory testing as the most suitable method for initial evaluation on field trials. Kinshuk (1996) provided a more detailed discussion of these evaluation techniques. Pham (1998) stated a brief review of other evaluation approaches such as objective-based, decision-based, value-based and naturalistic used in evaluating quality multimedia systems. However, only two of the approaches were found to be significant namely the naturalistic and value-based. Naturalistic approach is concerned with user's views, interests and experiences. The information for