Experimental and Multiple Methods Evaluation Models

These two models for designing and evaluating interactive multimedia products enable a mix of traditional and new methods of evaluation and are shown in the table below:

Summary of Two Evaluation Models					
Model & Description	**Paradigm**	**Where & How Used**	**Evaluation Methods**	**Pros**	**Cons**
1. Experimental evaluation model Suchman 1967 • Frequently used with control & experimental groups • Uses pre & post-tests & hypotheses	analytic-empirical-positivist-quantitative	In education & training sectors Suitable for summative evaluation	Quantitative pre-tests & post-tests	Scientific Tests specific questions Follows logical procedures	May miss important or unexpected outcomes of instruction Difficult to use randomised samples Tests not always reliable & valid No information about how teaching/learning intervention could be improved
2. Multiple methods evaluation model Mark & Shotland 1987 • Incorporates several techniques to evaluate complex programs • Triangulation to test adequacy • Bracketing to measure range	Eclectic-mixed methods-pragmatic	Web-based instruction Interactive learning Suitable for formative and summative evaluation	Quantitative & qualitative Eg Triangulation - questionnaire, interview, personality scale & impact on results Bracketing – range of responses to determine factors affecting learning eg motivation	Examines a wide range of features from various perspectives Prevents absolute certainty	Complex to design, implement, analyse and interpret Provides information about how teaching/learning intervention could be improved

Evaluation Models

1. Experimental (quasiexperimental) evaluation model

In an experimental design, cause-and-effect relationships are measured quantitatively through the study of experimental and control groups. A hypothesis would be tested, for example, 'the use of a structured Discussion Board and simulation enhance understanding of bioscience concepts and improve grades in a nursing/midwifery bioscience course ('Bioscience on the Internet)'. The experimental group would be given opportunities to discuss topics in different forums set up by the lecturer on the Discussion Board and given access to simulations, whereas the control group would have access to a Discussion Board but the discussions would not have pre-determined forum topics and simulations would not be used in the course.

In a true experimental design students would be selected randomly and then enrolled randomly into the course being evaluated and no other students would receive the benefits inherent in the course being evaluated, such as a structured Discussion Board. Ethically it might be difficult to use such a restrictive experimental design, using random selection, if the evaluator strongly believed that the learning strategy being studied improved student learning and performance. Hence the experimental design is modified to become quasiexperimental.

Experimental approaches are regarded as one of the most valid methods because they enable quantitative clarification that students have improved but the robustness and true empirical nature of the design may be compromised if modified ie becomes quasiexperimental. Both methods, experimental and quasiexperimental, are suitable for evaluating student learning about knowledge domains that include content that is predetermined, and there are clear learning outcomes, e.g., learning maths, chemistry etc. However, constructivist learning environments (where the achievement of learning outcomes are more difficult to quantify) could not be evaluated using these approachs as qualitative data is needed. For example, students could choose the group they wished to participate in, rather than being assigned randomly, as many might not want to try out new learning strategies or have their performance evaluated as closely. Hence those students who were prepared to put in the extra effort required by involvement in the experimental group might already be advantaged intellectually and be more motivated. Such variables can be

controlled for to a certain extent through the use of mechanisms such as pre-tests. Otherwise it would be difficult to measure the equivalence of groups.

A pre-test, taken before the program begins, is required for both experimental and quasi-experimental designs and controls for existing knowledge. Following the course a post-test is taken, and by comparing the scores between the tests the evaluator can determine whether students have made progress with their learning. Both the experimental and control groups would undertake this type of testing so that a comparison could be made between the two groups. In an evaluation of the course Biosicence on the Internet, the evaluator might see higher test scores for the students in the experimental group, but it would be difficult to test for an improvement in the understanding of bioscience concepts through the use of drill and practice type tests. Evidence of deeper, conceptual learning would require evidence of reflection and the use of problem-solving skills rather than higher scores in tests of memory. Also it would be difficult to pinpoint the use of the Discussion Board and simulation as directly responsible for the higher results especially if some students avoided using these multimedia tools. Hence the experimental model does not enable a broad evaluation of all factors contributing to the improvement in learning.

Also the range of variables likely to be present in a complex instructional design cannot be controlled for in a narrow experimental model, and a rigid, experimental evaluation design is not necessarily secure and predictive. If evaluators determined that students with particular learning styles gained higher results from using a structured DB but another group did not, and actually did better with another interactive strategy eg email feedback from the lecturer, then it would be more useful for instructors to obtain information on learning styles at the beginning of a class so that a range of learning strategies could be employed.

If the students really disliked using the DB but liked the simulation, these factors would be clearly stated in an experimental model, but the reasons would be more difficult to ascertain unless some form of qualitative analysis was used.

2. Multiple methods evaluation model

This model, which is really a set of guidelines developed by Mark & Shotland in 1987, is suitable for evaluating complex instructional programmes (use several levels of information and provide many opportunities for interactivity) because several methods can be used and both expected and unexpected outcomes can be measured. The model is particularly suitable for evaluating programmes with an interactive component (high-order learning) because responses can be measured at different levels and in different ways.

Two main approaches are used, triangulation and bracketing. In triangulation, one variable is measured by several different techniques. For example, improvement in learning through the use of a structured Discussion Board (DB) may be assessed via interviews, questionnaires (eg learning styles) and focus groups. Bracketing would examine a range of factors related to the improvement in learning. For example, a student may have stated in an interview that the use of structured forums in the Discussion Board did help her learning. Bracketing techniques would look at the factors which helped her learning eg structured DBs helped her organise her study, enabled contact with other students, motivated her to read about the topic. In other words a range of factors which led to improved results, in this case an improvement in learning, can be discovered and examined.

If multiple methods are used for evaluating programmes, researchers need to be aware of the need for skilled assistance when planning the methods to be used and statistical analysis. The multiple methods model generally means ideas taken from several areas rather than multiple quantitative techniques or multiple qualitative methods, but rather a range of methods which incorporate several ways of gathering and interpreting data.

Paradigms Associated with the Chosen Models

The experimental (quasiexperimental) evaluation model fits into an analytic-empirical-positivist-quantitative paradigm, one of the most established scientific schools of thought in the area of evaluation in education and the social sciences. Programmes to be evaluated can be reduced to parts and cause and effect between the parts established. Proponents in this paradigm believe that by keeping the factors (variables) to be measured separate from human beliefs and biases, objectivity can be maintained and mathematically 'true' data obtained.
The relationship between variables is analysed using statistics regarded as scientifically robust. Evaluation within this paradigm utilises controlled experimental techniques which measure whether one programme is better than another or one aspect of teaching and learning better than another. This paradigm is the mainstay for many academic institutions but alternative paradigms are being considered by more and more researchers as educational theory evolves and society changes.

The multiple methods evaluation model comes under the umbrella of the eclectic-mixed methods-pragmatic paradigm, one not traditionally used in educational evaluation, but probably the only paradigm capable of addressing the complexity of using technology in teaching and learning. Evaluation methods which are too simplistic will not enable important decisions to be made about the development, delivery and continuation of interactive instructional programmes. In this paradigm, several (eclectic) approaches eg quantitative, qualitative, can be used to examine information from different perspectives using different tools (mixed methods) eg questionnaire, document analysis, interviews, focus groups. Evaluators use appropriate tools within the relevant contexts to solve problems and make improvements (pragmatic) rather than predicting and controlling outcomes (empirical).

If the effectiveness of instructional design for interactive teaching and learning programmes is to be determined accurately and in a way which can signal strategies to improve a relatively new field of teaching and learning, then evaluation methods appropriate to the new technologies must be adopted. The eclectic-mixed methods-pragmatic paradigm and the multiple methods evaluation model have the most potential for the new environment of elearning. There is less likelihood of missing important indicators if a diverse and wide-ranging approach is taken, and the mistakes of 'tunnel vision' avoided.

Practical Activities Useful for Evaluating Interactive Multimedia Products

The purpose of the evaluation must be stated clearly before determining the model and the tools to be used in the evaluation process. If the only factor to be measured is whether students obtained higher exam grades by using either a structured DB or a simulation during the learning process, then an experimental model would be more appropriate and simpler to conduct than a multiple methods approach. If the evaluator wanted to explore whether a structured DB and simulation improved learning and exam grades and how and why the improvements occurred, then a more complex evaluation process, such as that found in the multiple methods model, is needed.

Mock Case Study

Lets assume that the use of a structured DB and a simulation of a bioscience concept such as fluid shifts in the human body improved the learning experience and outcomes for students taking an online course called 'Bioscience on the Internet'. These interactive learning systems would be evaluated differently depending on whether they were new innovations being tested in the course (formative) or established learning tools (summative). Formative evaluation is very useful during the developmental stages when designing interactive multimedia tools. Participants may be more open in their responses if they feel that the multimedia is being assessed rather than their achievements.
A formative evaluation design would look more closely at the learning process and learning environment rather than at learning outcomes eg grades. For the purposes of this assignment, a summative stance is taken when listing the evaluation activities. The means of determining improvement in learning is outlined in the next section using an experimental model and a multiple methods model. Evaluation of the learning process, learning outcome and appropriateness of the DB and simulation and student achievement within these areas become the focus.

Evaluation Activities in Experimental Model

Students enrolled in Bioscience on the Internet' were randomly assigned into two groups as follows:

Group A experimental - structured DB and access to a simulation package demonstrating fluid shifts in the human body.

Group B control - unstructured DB ie not arranged by topic and no access to the simulation package.

Both the experimental and control groups were given the same activities. The following data collection strategy is just one combination that could be adopted when employing an experimental approach.

- Survey to determine learning styles at the beginning of the course (provides information which may explain why some learners were more receptive to Discussion Board format for learning than others) .
- Confidence ratings of bioscience.
- Pre-test and post-test with general questions about bioscience and specific questions about fluids in the body.
- Statistical comparison of pre-test and post-test results between the two groups.
- Questionnaire – to determine usage of tools with closed questions eg 'did you use the DB?' 'did you access the simulation?' 'How many times did you use the DB or simulation?, The use of likert scales and ranking were included.
- Statistical analysis eg T-test to compare if there was a significant difference in exam grades for the two groups.
- Review of Discussion archives – frequency of use, type of discussion, depth of discussion.

Evaluation Activities in Multiple Methods Model

All students enrolled in Bioscience on the Internet' were treated as one group and given access to a structured DB and simulation package demonstrating fluid shifts in the human body. The following data collection strategy is just one combination that could be implemented.

- Survey to determine learning styles at the beginning of the course.
- Confidence ratings of bioscience prior to course and afterwards.
- Questionnaire – to determine usage of tools using closed and open questions to enable quick yes/no answers as well as descriptions of how multimedia was used or being used. For example, 'did you use the DB, if so how did it help your learning? If not, please explain. Likert scales and ranking questions were included to enable statistical analysis of usage of the DB and simulation. Hence quantitative and qualitative analysis was possible.
- Telephone interviews – to determine usage of the tools with individuals and in groups using teleconferencing. The interviews were taped and analysed.
- Analysis of Discussion archives - frequency of use, type of discussion, depth of discussions, language.
- Reflective journals examined for understanding and learning of bioscience concepts.
- Concept maps – evidence of linking key concepts.
- Critique of clinical scenarios assessed for evidence of problem solving and integration of bioscience concepts with clinical practice.

Triangulation Combinations

i. Survey, confidence ratings, questionnaire.
ii. Survey, confidence ratings, interviews.
iii. Discussion archives, reflective journals, concept maps.
iv. Survey, confidence ratings, critique of clinical scenarios.

One factor would be pinpointed as the reason for improved learning eg visual learners found the simulation useful.

Bracketing

This technique would take the information and present the range of responses within each of the activities. For example, all the factors influencing learning using the simulation and DB would be determined eg learning style, confidence with bioscience, time, existing skills with using the tools, access, preference for other tools for learning.

Conclusion

The difference between the two evaluation methods is that one aimed to provide information to "improve" the potential of intervention to foster better teaching and learning, whilst the other is aimed to "prove" whether the intervention did result in better learning. The multiple methods model enables more extensive evaluation tools to be utilised and would be less likely to miss factors not thought of by the evaluator than an experimental model. The evaluator is less interested in exam grades and the frequency of DB and simulation use and more interested in how and why the tools improved learning. The experimental model requires an experimental and control group and raises the ethical dilemma that one group was disadvantaged and missed out on the learning opportunities afforded by the DB and simulation. There is also the danger that results from a limited experimental model would be used predictively and determine future use of instructional technologies when only one or two aspects of learning had been examined.

A multiple methods model used within a eclectic-mixed methods-pragmatic paradigm gives more scope for exploration of the evaluation target, necessary in the complex and dynamic educational environment of our times.

References

Chomley, K. (2001) A comprehensive matrix on evaluation models. EDGI933/4 previous student work.

Frechtling, J. & Sharp, L. (Eds) (1997) User-Friendly Handbook for Mixed Method Evaluations. USA: Directorate for Education and Human Resources, Division of Research, Evaluation and Communication.

Gribbons, Barry & Herman, Joan (1997). True and quasi-experimental designs. *Practical Assessment, Research & Evaluation*, 5(14). Available online: http://ericae.net/pare/getvn.asp?v=5&n=14.

Learning Technology Dissemination Initiative. (1999) The Evaluation Cookbook. Heriot-Watt University, Edinburgh: http://www.icbl.hw.ac.uk/ltdi/cookbook/contents.html.

Phillips, R., Bain, J., McNaught, C., Rice, M, and Tripp, D. (2000) Handbook for Learning-centred Evaluation of Computer-facilitated Learning Projects in Higher Education. Murdoch University: Australia.

Reeves and Hedberg, J. (2001) Evaluating Interactive Learning Systems. Englewood Cliffs NJ: Educational Technology Publications.

Wilde, J. & Sockey, S. (1995) Evaluation Handbook. New Mexico Highlands University: Albuquerque, NM.