# SORTING THE WHEAT FROM THE CHAFF: EVALUATING A LARGE CORPUS OF COMPUTER-BASED STATISTICS TEACHING RESOURCES

Erica J. Morris, Institute of Educational Technology. The Open University
Martin Le Voi, Faculty of Social Sciences. The Open University

*A project was conducted to select computer-based statistics courseware for a psychology masters course in advanced experimental design and analysis. This involved a survey of computer and web-based resources for statistical teaching followed by a detailed evaluation of selected resources. Sixty-nine resources for statistical teaching were initially evaluated and eleven of these resources were selected for further detailed investigation. For this, an evaluative framework that provides assessment for the usability and possible instructional effectiveness of a computer-based learning resource was used. Five of the eleven resources were considered worthy of further detailed examination, including an extensive testing procedure with students. A variety of issues concerning the evaluation of computer-based resources for statistics teaching are discussed.*

## THE AIM OF THE PROJECT

The aim of the project was to survey, evaluate and select computer-based statistics resources for a psychology masters course in advanced experimental design and analysis. This paper describes the first phase of the project which involved:

- A survey of computer-based resources for statistics teaching
- Preliminary evaluation of statistical resources
- Expert formal evaluation of selected resources

## THE EVALUATION OF COMPUTER-BASED TEACHING RESOURCES

Statistics is taught to a large range of students at many levels, and often is seen as "hard" by students e.g. in psychology, for whom it is not the main focus of interest. Being largely quantitative, and with some aspects of it having explicit geometrical interpretations, it has attracted many attempts to improve the student's learning experience by using computer and/or web based teaching. With so many potentially useful resources already in existence, there seems no need to "re-invent the wheel" and produce new teaching resources if perfectly good ones are already around. The problem, therefore, is finding and selecting them. Selecting usable resources is likely to be especially difficult, since, if there are many, student-based evaluation will be impractical, although this is frequently recommended (e.g. Gill, Dick, Reiser and Zahner, 1992; Reiser and Kegelmann, 1994). On the other hand, uninformed unsystematic subjective reviews

of resources may be quick but inefficient at identifying the most useful resources. What is required is an efficient formal system which allows systematic evaluation of large numbers of resources.

Expert evaluation that makes use of formal criteria has been used to assess the quality of computer-assisted learning programs (Bangert-Drowns and Kozma, 1989; Shuell and Schueckler, 1989). These approaches were aimed at the evaluation of computer-assisted learning programs, but may be extended to web-based teaching resources. Therefore, a framework, which is based on previous research (Bangert-Drowns and Kozma, 1989), was used for the first phase of the project.

## A SURVEY OF STATISTICAL RESOURCES

To find existing resources for statistics teaching a survey of the World Wide Web was conducted and directories of software were consulted.

A total of sixty nine statistical resources were surveyed (fifty computer-assisted learning programs that included multimedia resources, eighteen web-based resources and a glossary).

### *The evaluative framework*

The evaluative framework was based on criteria described by Bangert-Downs and Kozma (1989). These criteria were chosen because they provide assessment for the usability and the possible instructional effectiveness of a computer-based learning resource and also because they are informed by research and theory in instructional design. The developed framework consists of selection, content, operational and instructional criteria (Table 1). The selection criteria were used in the preliminary evaluation of statistical resources for teaching.

### *Preliminary evaluation of statistical resources*

A preliminary evaluation was conducted to determine if particular resources were suitable for the masters course. More specifically, the resource had to satisfy four *selection* criteria (see Table 1).

From the preliminary evaluation, eleven resources were selected for further

detailed evaluation (nine computer-assisted learning programs and two web-based resources).

*Evaluation of selected resources*

Each part of the framework consists of a set of relevant evaluative criteria (Table 1). Each of the eleven resources were evaluated with respect to these criteria. This provided a qualitative assessment of particular aspects or features of the resource.

To demonstrate the application of the criteria to the selected resources, the *abridged* evaluation of a computer-assisted learning program is described. Further details of the evaluation of statistical resources for teaching is available in Morris and Le Voi (1997).

*Content criteria*

In ConStatS there is accurate coverage of the statistical topics. It was difficult to judge how thorough the treatment of the topics covered in the resource are because no specific preinstructional objectives are presented.

*Operational criteria*

The general speed of the program is adequate. The user manual for ConStatS is comprehensive where it includes details of the rationale for the development of the resource, guides for working through the different sections of the resource and a glossary. ConStatS' user interface is not particularly appealing or pleasant to look at. The ConStatS environment is inconsistent where, for example, the style of dialogue boxes and font change within the program. Furthermore, the ease of navigation within the program is poor. The user interface provides no indication of the structure or relative size of the program and its topic sections. In addition, the program makes inappropriate use of the command exit. Buttons entitled exit are provided throughout the program. At times, this means that two exit buttons can be on the screen at once, but one will let the user quit the program, and the other will close the section the user is currently working on.

Table 1. Evaluative framework. (Based on criteria provided by Bangert-Drowns and

Kozma, 1989).

| | |
|---|---|
| 1. Selection criteria | 1. 1 Are the target users of the resource psychology students?<br>1. 2 Are the statistical topics relevant to the target course?<br>1. 3. Is the resource compatible with MS–DOS or Windows 3.11?<br>1. 4. Category of resource |
| 2. Content criteria | 2. 1 Accuracy<br>2. 2 Thoroughness of treatment |
| 3. Operational criteria | 3. 1 Speed of execution<br>3. 2 Quality of documentation<br>3. 3 Treatment of operational errors<br>3. 4 Quality of user interface<br>3. 5 Ease of navigation |
| 4. Instructional criteria | 4. 1 Presence and quality of preinstructional introduction<br>4. 2 Degree of learner control<br>4. 3 Frequency and variety of practice exercises<br>4. 4 Quality of feedback<br>4. 5 Control of feedback availability<br>4. 6 Use of a variety of representations (text, graphics, animation)<br>4. 7 Use of sound<br>4. 8 Motivational quality<br>4. 9 Clarity of learner options (simulation software)<br>4. 10 Clarity of decision effects (simulation software)<br>4. 11 Availability of record of results (simulation software)<br>4. 12 Use and availability of student record (e.g. progress, performance)<br>4. 13 Co-ordination of components |

*Evaluation of ConStatS: Software for Conceptualising Statistics*

*Instructional criteria*

ConStatS offers no preinstructional introduction or learning objectives. There are no practice exercises in the program and although this restricts feedback, appropriate program feedback is invoked in response to a user's actions at the interface. The resource does not use a variety of representations. In the main, only text and graphs are used and the resource does not provide any audio.

In the three sections that comprise the Sampling part of the program, the clarity of the learner options and the decision effects for the simulation is poor. For example, whilst a simulation of sampling from a population is taking place on the screen, there are too many small windows and graphs displayed with no sound or text-based commentary available to give the learner a clear idea of the process that is being illustrated. For the simulations, the learner cannot keep a record of their on-screen experiments. The program does not appear to have any kind of student record that details, for example, those particular sections the learner has covered.

*Conclusion*

ConStatS is a comprehensive computer-assisted learning resource which includes a detailed manual. The user interface of this resource is, however, unappealing and is not consistent. The resource lacks necessary instructional objectives and it does not provide structured tasks or exercises at the user interface for the learner to work through. This evaluation of ConStatS suggested that the resource should not be used for the masters' courseware.

## FURTHER RESEARCH

After the evaluation, five of the eleven resources were considered worthy of further detailed examination. This will lead to the second phase of the project, which will involve students as participants in the evaluation process, and will therefore include extensive user testing of the selected computer-based learning materials.

## DISCUSSION

This evaluation highlighted several issues that concern the evaluation of computer-based resources for statistics teaching. Different user disciplines tend to focus on particular statistical topics and in the selection criteria used in the project the aim was to evaluate software that was targeted at psychology students. It was a surprise to find that

some resources did not make explicit the anticipated users of the resource. Concomitantly, it is clear that resources are aimed at students from a variety of user disciplines of statistics. Moreover, some of the computer-based resources lacked the necessary documentation to make optimum use of their facilities.

Computer-assisted learning programs in particular, are sometimes linked to certain data-analysis software. This is problematic: an institution may already use and support a data-analysis package, but the computer-assisted learning program which makes use of a different package cannot therefore be easily integrated into the curriculum. This kind of design decision limits the contexts in which a computer-based resource can be used in higher education.

A formal evaluative framework based on previous research (Bangert-Drowns and Kozma, 1989) was used to provide a qualitative assessment of computer-based resources for statistics teaching. Previous research has assessed instructional software on these criteria, but has also made use of experts ratings on 5-choice evaluative items (Bangert-Drowns and Kozma, 1989). The present framework could be modified in this respect, where an evaluator's rating of between 1 and 5 is given for an evaluative item that concerns, for example, accuracy of content.

This evaluative framework could be used whenever computer-based resources for statistics teaching must be selected, but when extensive evaluation involving students is not feasible in the first instance.

There are other approaches to evaluating computer-based resources, such as GOMS (John and Kieras 1996) and CCT (Kieras and Polson, 1985). These approaches are task-oriented and require a heavy analytical load. They are also focused mainly on user interface analyses. More recently, the Ontological Sketch Model (Blandford and Green, 1997) has attempted to reduce the skill required of the analyst, but still requires substantial analytical work and is also focused towards usability analysis. The approach we have developed here attempts to make explicit the formal criteria being used to evaluate the various resources, without making the whole procedure so labour-intensive that it cannot in practice be applied to a large number of resources, which was the design goal of the project.

It is, however, recognised that the outcome of a formal evaluation must be followed by evaluation studies that involve extensive user testing. The evaluation of computer-based resources must be empirical where both quantitative and qualitative data

is gathered concerning students' use of the resource (Draper, Brown, Henderson and McAteer, 1996). Such an evaluation will form the final stage of this research.

REFERENCES

Bangert-Drowns, R. L. and Kozma, R. B. (1989). Assessing the Design of Instructional Software. *Journal of Research on Computing in Education, 21(3)*, 241 - 262.

Blandford, A. and Green, T. R. G. (1997). OSM: an ontology-based approach to usability evaluation. Paper presented at the Workshop on Representations, Queen Mary College, London. Also available via http://www.ndirect.co.uk/~thomas.green/workStuff/res-OSM.html.

Draper, S. W., Brown, M. I., Henderson, F. P., and McAteer, E. (1996). Integrative Evaluation: An Emerging Role for Classroom Studies of CAL. *Computers and Education, 26(1-3)*, 17 - 32.

Gill, B. J., Dick, W., Reiser, R. A., and Zahner, J. E. (1992). A New Model for Evaluating Instructional Software. *Educational Technology, 32(3)*, 39 - 44.

John, B. and Kieras, D. E. (1996). Using GOMS for user interface design and evaluation: which technique? *ACM ToCHI*. 1-30.

Kieras, D. E. and Polson, P. (1985). An approach to the formal analysis of user complexity. *Int. J. Man-Machine Studies, 22*, 356-394.

Morris, E. J. and Le Voi, M. (1997). Using computers and web-based statistical resources. *Department of Psychology. Report Series*. The Open University. Access via http://socsci.open.ac.uk/OTD30/report.

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

Shuell, T. J. and Schueckler, L. M. (1989). Toward evaluating software according to principles of learning and teaching. *Journal of Educational Computing Research, 5(2)*, 135 - 149.