

The evaluation of computer-aided learning in medicine

Computer-aided learning (CAL) has been proposed as a panacea for all educational problems: it utilises an efficient, learner-centred pedagogy, and it may be a least-cost solution compared to the alternatives. Whilst CAL can never be a complete panacea, it would appear able to provide at least part of a solution to the demands of recent changes in medical education, in particular the shorter, more structured postgraduate training and the new core curricula at undergraduate level.

However, these potential benefits may not be realised because CAL usage is not sufficiently widespread, especially in postgraduate medical education. Much of the reluctance to use or produce CAL may be because of the perceived paucity of clinically acceptable evaluation data. Clinicians routinely use a different research paradigm to educationalists and the randomised controlled trial, the gold standard of clinical research, is often not applicable to educational research. There is a requirement for an increased awareness from clinicians of other, more qualitative research for CAL evaluation.

In recent years there has been a rejection of the controlled experimental design of CAL evaluation outside medical specialties (Isroff *et al*, paper presented at WICC'95 in Birmingham, UK; Draper, report on ELTHE 3 workshop www.psy.gla.ac.uk/~steve/ELTHE.report.htm).^{1–3} Laurillard has highlighted the discrepancy between the questions asked of evaluation studies in new technology, and the answers they provide. Politicians, decision-makers and funding bodies, consistently ask the question “do learning technologies improve learning?” and the evaluation studies consistently avoid this issue and demonstrate instead that “learning technologies have the potential to improve learning”. She makes the statement that “there is no way past this impasse”. “The project demonstrated clear potential for the medium to be educationally effective” is a common finding in evaluation studies of new technology and should no longer be regarded as news. Just as there never seems to have been a need to evaluate the impact of a lecture or textbook in isolation from all other supporting resources, so there appears to be little justification for doing the same with CAL; a more holistic approach to evaluation is required.

Evaluation of CAL in medicine is in its infancy and results of many studies will hopefully lead us to find out how it can best be utilised. Some learning tasks may be more suitable for CAL than others and we need to find out which subjects are amenable to CAL and which are not, rather than adopt technology simply because it is available. CAL may be used by individuals or groups; groups may function independently or may be facilitated by a tutor who adopts a CAL package as a resource for teaching just as they may traditionally use print-based materials.

Medi-CAL

There have been a number of studies that support the use of CAL materials in medicine. A meta-analysis of 49 trials compared CAL and traditional methods, examining cognitive outcome, behaviour and learning theory, as used by a variety of medical trainees, students and doctors. In 98% of the trials, the performance of the CAL students was at least as good as conventional means and in 61% of trials, performance improved on CAL, as measured by pre- and post-test scores. In the majority of cases the times taken to achieve the same score were the same or decreased in the CAL groups. Although in some studies more time was

taken by the CAL groups, this was ascribed to either enhanced motivation or increased opportunity.⁵

Gathy *et al*⁶ demonstrated that weaker students improved more than stronger students using the same CAL tutorial, and that all the students expressed a preference for the computer-based material. It would have been interesting to ask why they expressed this preference. Another study demonstrated that a CAL tutorial on the early diagnosis of substance abuse was at least as good as experiencing the ‘real thing’, and was significantly superior to other conventional teaching methods,⁷ while a recent study from America demonstrated that a group of university sociology students who did all of their learning through the Internet did better than those taught conventionally.⁸

CAL materials can be used to reach a wider audience than conventional means, and this contributes to the perceived cost-effectiveness. The introduction of CAL reduced traditional ‘chalk and talk’ lectures by one-third in one veterinary school.⁹ However, CAL is not a cure-all. The latest technology can be mesmerising and there is a need to ensure that the production of CAL materials is driven by educational goals rather than by what is available. The Royal College of Surgeons have deliberately eschewed computers for their excellent distance learning programme.¹⁰ They felt that trainees would prefer to use more familiar media, such as paper-based text and pictures, video and audio tapes, although there are unfortunately no data to support this supposition. This approach also circumvents problems with hardware compatibility, computer literacy, etc.

Traditional approaches to CAL evaluation

A common approach to evaluation found in the medical literature is to apply the standard medical paradigm of quantitative research and directly compare CAL with ‘traditional’ teaching and learning situations. Many published studies describe selecting two groups, one using some sort of new CAL resource and the other acting as the control, using pre- and post-testing or examination results.^{11–16} CAL packages have been shown by such studies to be useful. However this approach is too simplistic. For example, there are no gold standards for other media. Also, finding and randomising comparable groups, and controlling all the possible conflicting variables that have an effect on learning is difficult, if not impossible.

The concept of ‘quality’ in educational software, ie, what makes a piece of software a good teaching or learning resource, is also complex. As previously noted, this may be because of the interaction of a large number of uncontrollable variables, which are complex and frequently difficult to define. The Centre for Educational Research and Innovation¹⁷ identified a number of potentially important factors, eg, the subject matter and the curriculum, the teaching and the teacher, the learning and the learner, the technology itself, and the design of the software. All of these interact synergistically, to varying degrees, to make a piece of software effective or ineffective as a teaching and learning resource. Learners themselves have different amounts of pre-existing knowledge and different needs. Learning is a social activity and in the real world we obtain information from many sources: books, journals, colleagues and patients, and it is impossible to control for these factors in a pseudoscientific ‘laboratory’ experiment.

There are also difficulties in interpreting the results of pre/post knowledge questionnaires. Most studies focus on

the immediate results of an objective assessment of an increase in knowledge, whereas there may be delayed learning gains which are more important. Longitudinal studies demonstrating a change in actual behaviour in a particular clinical situation might be a more important.

Studying the examination results of those exposed to a particular intervention and comparing this group to those who did not have the intervention carries a risk when drawing conclusions as to the value of the intervention. Factors such as auto-compensation, where learners find one resource unsatisfactory and compensate by seeking an alternative, should be considered.

Measurement of attitudes to computers may also be an important factor. Draper *et al*¹⁸ reported that the expectations (both positive and negative) of the learner and the teacher have a great influence on attitudes to a particular CAL package. Computer enthusiasts tended to be more positive when using a package for the first time than those with little computer experience. It was only with repeated use of a package that the content of the CAL rather than the delivery platform itself became a factor. Finally, all researchers need to be aware of the Hawthorne effect (performance is enhanced by being part of an experiment).

Why quantitative methods do not apply to educational research

Qualitative research methods have a long history in the social sciences but medical research is fixed upon the scientific paradigm of research methodology with the gold standard of experimental design being the randomised double-blinded trial. However, this approach cannot be directly transferred to educational research; we need to examine cultural aspects, and try to unravel the *why* and *how*? Classically, evaluation is divided into formative and summative. More recently, alternative types of evaluation have been described: illuminative evaluation, integrative, and situated evaluation. Formative evaluation is used to improve the design of the software through iteration. Summative evaluation is used after software design and production to establish performance and properties. Illuminative evaluation is an observational approach, used to discover what factors and issues are important in actual practice in effecting an outcome

New approaches to CAL evaluation

Beeby²¹ defined educational evaluation as a process whereby information is gathered from many different sources, interpreted, and conclusions drawn by a process of triangulation. An evaluation or review is a judgement about the value, appropriateness and usefulness of a piece of educational software. In the UK, a software evaluation is regarded as a critical appraisal of software involving in-depth testing using appropriate quantitative or qualitative research methodologies; anything else is a software review. However, in the US and Australia, the two terms are synonymous so that an evaluation is a broader term that includes critiques by academics and sales descriptions (Centre for Educational Research and Innovation, 1989). The difficulties of evaluating a piece of educational software are compounded by national or cultural differences which can result in a piece of software considered excellent in one country, being judged as of little value in another.¹⁷ Real world use is often deemed the most important factor in any CAL evaluation. Draper *et al*²² reject the idea of summative evaluation in favour of the term integrative evaluation, ie, how it is used and what are the outcomes in the teaching and learning situation. Integrative evaluation can help teachers make better use of CAL by adjust-

ing how it is used, rather than by changing the teaching programme, which may be expensive. This is particularly pertinent to postgraduate medical education, where there is often no single correct answer. Teachers need to be aware that this type of CAL is likely to be unsuitable for individual use as the ambiguity of the answers may create confusion; it may therefore be more appropriate for group learning, and may be particularly useful for fostering critical thinking. Draper *et al*²⁰ also recognised that the performance of CAL may depend on the context of use and they advocated 'real' observation, in combination with measurement of learning gains using confidence logs and quizzes to form an *integrative* evaluation. Participants in the ELTHE workshop also advocated looking for unspecified educational gains.

Gunn²¹ describes an evaluation framework, which has been developed specifically for CAL research. Situated Evaluation of CAL (SECAL) is an evaluation in an authentic context, ie, looking at the combined effects of CAL, other resources, aspects of the learning environment and qualitative case-based research. SECAL requires examination of the 'bigger picture' to include institutional factors which might influence culture, attitudes and acceptance of new ways of teaching and learning

The Open University probably has the most experience in the use of computers in distance education in the UK and recently published a report of the first 15 years of evaluating CAL. They have found that a combination of quantitative and qualitative methodologies yields the most information. Questionnaires (consisting of multiple-choice and open-ended questions) are combined with telephone and face-to-face interviews. Although learner diaries and journals did provide some useful information, they were of less value as most students did not complete them. The main barriers to computers in teaching and the learning environment were the lack of privacy, the fear of looking stupid, and the fear of breaking the hardware.³

How to evaluate CAL in medicine

It needs to be emphasised that CAL evaluation should not just examine learning outcomes: researchers need to examine the bigger picture, to look at cultural aspects and attitudes to computers in teaching and learning. Moreover, it is essential to determine whether potential users have access to suitable hardware and the requisite skills. Ideally, CAL evaluation results would be transferable and not just relevant to enthusiasts and developers. The most useful outcome measure is therefore probably whether a package is actually used and/or whether learners find it useful in the real world, rather than whether an increase in learning can be demonstrated. In this context, market forces may determine its value.

Quantitative methods aim for reliability (consistence on re-testing), and qualitative methods score more highly on validity. We advocate a range of methods. Qualitative methods include observation, in depth interviews, focus groups and case studies. Focus groups are particularly suited to the study of attitudes and experiences.²⁴ Educational studies often deliberately select, ie, purposive sampling (deliberate choice of respondents) rather than statistical sampling. Triangulation, use of different research methods in combination, can be a very useful check of validity.

Surveys can also produce quantitative information. They can provide the *how many?* type answers but not the *why?* Pre- and post-testing does have a role but should be seen as just one tool in the whole picture of the evaluation. Confidence logs have proved surprisingly useful as a tool by many evaluators outside medicine.¹⁸ They can be used pre-

and post- or during an experimental session and can provide familiar quantitative results suitable for statistical analysis. Focus groups can be used to aid questionnaire design. They provide useful information after a questionnaire has been performed as they may shed light onto why something in a questionnaire has scored well or badly. Focus groups may also be used to concentrate on particular areas of interest identified by other methods. Observation of learners using a CAL package may also be helpful. Difficulties in using the program can be identified (with recommendations made to alter software, produce a printed instruction or offer computer skills training). In-built audit trails can achieve a similar end.

Conclusion

CAL evaluation is complex and investigators cannot rely solely on one research paradigm. It is important that the limitations of experimental design are recognised. Evaluators need to examine the use of CAL in the 'real world'.

Evaluation should include consideration of cultural aspects, and explore the teaching and learning environment. Evaluation must be pragmatic, not dogmatic. In particular, qualitative tools from the social science fields, such as questionnaires, focus groups, case studies, interviewing techniques and observational studies may be useful, and can be combined with more familiar research methods in medical research. Hopefully this will enable medical educators to exploit all the potential benefits afforded by CAL.

JANE PANIKKAR

Department of Obstetrics & Gynaecology, Northern General Hospital, Sheffield, UK

TIM DRAYCOTT

Department of Obstetrics & Gynaecology, Gloucestershire Royal Hospital, Gloucester, UK

JULIAN COOK

Learning Technologies Evaluation, Institute of Learning & Research Technology, University of Bristol, Bristol, UK

Accepted 23 April 1998

Keywords: computer-aided learning; education

- 1 Parlett M, Dearden G. Introduction to illuminative evaluation. *Studies in higher education*. Pacific Soundings Press, 1977.
- 2 Gunn C. Future directions for CAL evaluation. (abstract). ALT-C96 conference programme and abstracts. Glasgow, Scotland, 1996; p108.
- 3 Jones A, Scanlon E, Tosunoglu C, et al. Evaluating CAL at The Open University: 15 years on. *Comput Educ* 1996;26:5-15.
- 4 Laurillard D. What's the problem, Mr Wolf? *Times Higher Educ Suppl* 1993;22 June.
- 5 Kulick CC, Kulick JA. Effectiveness of computer-based instruction: an updated analysis. *Comput Hum Behav* 1991;7:75-94.
- 6 Gathy P, Denef J-F, Haumont S. Computer assisted self assessment in histology. *Comput Educ* 1991;17:109-16.
- 7 Brown RL, Carlson BL. Early diagnosis of substance abuse: evaluation of a course of computer assisted instruction. *Med Educ* 1990;24:438-46.
- 8 Brake D. The Net benefits of learning from home. *New Sci* 1997;2067:5.
- 9 Garvin AP, Carrington SD. The production of hypermedia peer tutorials by undergraduate veterinary students as assessed by 'Directed self-education' group projects. *Conference proceedings: Computers in Medical Education 1995*. Bristol: CTI Centre for Medicine, 1995; pp 19-24.
- 10 Cochrane J. Distance learning courses from the Royal College of Surgeons. *Br J Hosp Med* 1997;270-1.
- 11 Parish RA. The efficacy of CAI in an outpatient paediatrics clinic. *Clin Paediatr* 1986;25:255-6.
- 12 Papa FJ, Meyer S. A computer assisted learning tool designed to improve clinical problem solving skills. *Ann Emerg Med* 1989;18:269-73.
- 13 Rasheed KS, Cohen PA. An evaluation of computer based instruction versus printed study guides in a dental material course. *J Dent Hyg* 1990;64:36-9.
- 14 Porter RS. Efficacy of CAI in the continuing education of paramedics. *Ann Emerg Med* 1991;20:380-4.
- 15 Halloran L. A comparison of two methods of teaching. Computer managed instruction and keypad questions versus traditional classroom lecture. *Comput Nursing* 1995;13:285-8.
- 16 Elves A, Ahmed M. Urology computer assisted learning programmes for medical students. (abstracts). ALT-C96 conference programme and abstracts. Glasgow, Scotland, 1996; p86.
- 17 Centre for Educational Research and Innovation. *Information technologies in education. The quest for quality software*. Paris: OECD, 1989.
- 18 Draper SW, Brown MI, Edgerton E, et al. Observing and measuring the performance of educational technology. A report by the University of Glasgow's institutional project in the Teaching and Learning Technology Programme, *Teaching with Independent Learning Technology*. Glasgow: University of Glasgow, 1994.
- 19 Beeby CE. *Curr Issues Educ* 1977;4:67-78.
- 20 Draper SW, Brown MI, Henderson FP, McAteer E. Integrative evaluation: An emerging role for classroom studies of CAL. *Comput Educ* 1996;26:17-32.
- 21 Gunn C. CAL evaluation: what questions are being answered? *Comput Educ* 1996;27:157-60.
- 22 Kitzinger J. Introducing focus groups. *BMJ* 1995;311:299-302.