Research Paper ■

Systematic Assessment of World Wide Web Materials for Medical Education:

Online, Cooperative Peer Review

ELIZABETH BERRY, PhD, CHRISTINE PARKER-JONES, MSc, RICHARD G. JONES, DM,
PATRICK J. R. HARKIN, MBChB, HAROLD O. HORSFALL, MBChB,
JOSEPH A. NICHOLLS, PhD, NICHOLAS J. A. COOK, MSc

A b s t r a c t   **Objective:** To develop a generic methodology for the online assessment of medical education materials available on the World Wide Web and to implement it for pilot subject areas.

Design: An online questionnaire was developed, based on an existing scheme for computer-based learning material. It was extended to involve five stages, covering general suitability, local suitability, the user interface, educational style, and a general review. It is available on the Web, so expert reviewers may be recruited from outside the home institution. The methodology was piloted in three subject areas—clinical chemistry, radiology, and medical physics—concentrating initially on undergraduate teaching.

Measurements: The contents of completed questionnaires were stored in an offline database. Selected fields, likely to be of use to students and educators searching for material, were input into an online database.

Results: The online assessment was used successfully in clinical chemistry and medical physics but less well in radiology. Fewer resources were found to fit local needs than expected.

Conclusion: The methodology was found to work well for topics where teaching is highly structured and formal and is potentially applicable in other such disciplines. The approach produces more structured and applicable lists of resources than can be obtained from search engines.

■ **JAMIA.** 1998;5:382–389.

Affiliations of the authors: The University of Leeds, Leeds, England (EB, CP-J, RGJ, PJRH, NJAC); St. James's and Seacroft NHS Trust, Leeds (HOH); University of Wales College of Medicine, Cardiff, Wales (JAN).

Correspondence and reprints: Dr. Elizabeth Berry, Medical Physics, University of Leeds, Wellcome Wing, Leeds General Infirmary, Leeds LS1 3EX, United Kingdom. e-mail: ⟨eberry@leeds.ac.uk⟩.

There is a growing supply of medical resources for teaching and learning available on the World Wide Web. Suitable material could be integrated into undergraduate medical curricula to complement existing teaching methods, by providing additional reference and support materials for lectures, tutorials, and other modes of teaching.[1,2] Theoretically, materials on the Web should be of value to educators around the world, but there are several reasons why they may not be. For example, the resources are unlikely to have been subject to external assessment for quality; their continued availability is uncertain; they may be too advanced for a medical student new to a subject; and they will not have been designed to fit the curriculum as structured in the home institution. Although the Web may be considered analogous to a university library, a number of fundamental features apply only to libraries and not to the Web. For example, books

and journals are pre-selected by librarians and lecturers to include material that is known to be of relevance to local courses. The selection of new books is based on descriptions in publishers' catalogues, on reviews in journals, and on word of mouth, while subscriptions to journals are taken up according to the subject area and reputation of each journal. Books undergo a lengthy editorial process, and before purchase most books are read, or at least perused, for content and quality. Articles in a journal are peer-reviewed before inclusion. Within libraries there are a well-established cataloging system and numerous mechanisms for locating information both within the library and in the wider external context of nationally available collections.

Thus, information on the Web differs in several ways. There is little or no peer review, editorial input, or pre-selection.[3] Cataloging on the Web is in its infancy,[4] and although there are multiple mechanisms for locating material, search engines indiscriminately return large unstructured lists of information not particularly relevant to the user's needs.[5] Naive users have difficulty refining their search, which is critical for focusing the response. These issues must be addressed if the Web is to be used to its true potential. The task has been begun by organizations such as OMNI[6] and Health on the Net,[7] which are undertaking reviews and indexing medical Web resources; their reviews, however, are not specifically for teaching materials, nor can they consider how specific material fits any local requirements.

Inevitably, students will wish to explore the Web in search of additional material, but given the sheer size of the resource they will waste a vast amount of time looking at incorrect or inappropriate material. The objective of this pilot study was to develop a methodology for assessing Web material for educational use, initially for undergraduates but generic enough to suit postgraduates, too. Criteria for assessing educational materials have been developed, and an online methodology for cooperative peer review is described.

Background

Once the provision of information via the Web became popular, the need to review and index resources quickly became apparent. Information specialists in the medical field, including OMNI,[6] were the first to provide online indexing, with short descriptions designed to help potential readers find appropriate material. For teaching materials, however, the information provided is generally inadequate as the basis for an informed decision about utility. It is emphasized that the type of review required, which is essentially

intended to help the educator select appropriate resources, is quite separate from the evaluation of the effect of an educational package on learning outcomes.[8] Information to help educators make decisions about the usefulness of resources is likely to be published in periodicals (such as CTICM Update[9]), but these reviews may be highly individual in style and lack a formal methodologic structure. Descriptions of more structured approaches appear in the literature; for example, an early tool designed to facilitate the decision-making process regarding both audiovisual and computer programs in nursing was described by Van Ort.[10] Posel[11] covered similar guidelines written for hospital-based nurses. More recently, a structured approach to reviewing computer-based material was described by Huber and Giuse,[12] who used a bi-level evaluation process to ensure that both educators and information professionals were able to contribute to the overall review. The part of the process performed by the educator was deliberately kept short, concentrating on usability, and omits consideration of educational methodology[13] and accuracy of information.[14,15] A consumer-oriented model for advising purchasers has been developed,[16] which is designed to direct educators to the programs that suit their needs. The special case of delivery over the Internet was tackled in a software performance evaluation[17] where the authors concentrated on a single performance metric characterizing response time.

The amount of material available on the Web is increasing exponentially, and advice for potential users is needed if they are to make decisions about integrating resources. Although the studies described above tackle some of the important issues, none is comprehensive. For example, although much of the teaching material available is of high quality, a significant proportion has been created without reference to formal educational methodologies or consideration for the problems of delivery within other curricula or hardware configurations. Similarly, if the materials under review are to be used by individuals studying a particular course or following a given curriculum, it is important to know how well the material will support this. For speed and convenience the review process should allow an early exit at predetermined stages if material is unsuitable. As it must also be simple to perform and easily accessible to reviewers, the review process described in this paper has been implemented on the Web. An online process has the additional advantage of allowing a reviewer to begin work immediately without postal delays or administrative overhead and assumes nothing about the reviewer's hardware and software environment and skills. A reviewer may also have immediate access to the latest reviews by others.

Table 1 ■

Quality Criteria for Computer-based Learning Medical Education Materials

Is the information contained in the material:

- Correct?
- Up to date?
- At an appropriate level for the intended course(s)?
- Directly mappable onto identifiable section(s) of the course Aims and Objectives?

Does the package run:

- Under Microsoft Windows, any version?
- Under DOS with a graphical interface?

Does the software:

- Require more than 8 Mb?
- run on Novell 4 networks?
- Require video-disk, touch-screen, digitizing tablet, scanner, or other dedicatesd hardware?
- Require CD-ROM drive?
- Require a soundcard?

Methods

Development of Quality Criteria for CBL

The assessment criteria used for the review of Web materials were derived from less specific guidelines the authors have previously used in the assessment of traditional computer-based learning material.[18] These guidelines were developed to enable reviewers to process material quickly, allowing rapid rejection of material that is likely to be unsuitable. The guidelines were based on the recommendations of the U.K. National Education initiatives Computers in Teaching Initiative (CTI)[19] and TLTP Teaching and Learning Technology Programme (TLTP),[20] together with lessons learned from practical experience. In order for this questionnaire to be useful, it must be simple and quick to use. Our original questionnaire asks reviewers to consider whether material meets a set of simple guidelines (Table 1). The early questions address educational issues; later questions address specific local technical limitations, and a final section (not shown in the table) addresses legal and financial implications of implementing the package. A single negative response in the first section is considered sufficient to grade an item as unsuitable for our purposes, and the reviewer may abandon the process at this stage. Later sections require more consideration; reviewers are given guidance on how to interpret the answers.

The guideline questions are organized in a hierarchic format that maximizes the use of reviewer time by making it possible to reject highly unsuitable material at an early stage. Reviewers using these criteria can therefore deal with more material than would be expected by quickly transferring their attention from material likely to prove unsuitable and can devote more time to the fuller assessment of more complex material. The amended criteria for Web materials are described later.

Transfer to Online Format

The online questionnaire was implemented for the Leeds Interactive Medical Education (LIME) project using HTML forms.[21] A reviewer calls the questionnaire from a list of resources available for assessment by selecting a ''Review'' option. A form—with fields relating to the title and URL of the resource already completed—is returned to the user's browser, and a second browser window containing the resource is opened. This is accomplished using a common gateway interface (CGI) script, which allows processing of data by programs on our server (Figure 1). The reviewer completes the form and presses the ''Send'' button; another CGI script ensures that an acknowledgment is returned to the reviewer.

A Web-based database is used to store the results of reviews. For the pilot project, a simple flat-file text-based database was developed to which data entry may be made manually via a Web page. A reduced set of fields was used, with priority given to those to be used by students and educators searching for useful material. All the reviewers' comments are saved separately in an offline database.

Stage 1: Preliminary Assessment

The online review of resources is carried out in three parts. The design aims to minimize the input required by reviewers by allowing the review to be stopped at one of three points during the process.

A preliminary assessment of materials is performed based on very broad selection criteria. This stage is designed to exclude inappropriate material from further review, for example, syllabuses are often available on the Web and are returned by automatic searches. It is done without domain expertise and ensures that the material is in one of the educational categories suggested by Laurillard[22] and listed below:

- Tutorial
  Modified Socratic dialogue
  Drill and repeat
  Simulation

Free-form or directed exploration

- Case studies with feedback
- Textbook
- Lecture slides
- Self-assessment with answers or feedback

It is also determined at this stage whether the resources will run on the computer facilities available to students at the home institution.

Stage 2: Resource Assessment

The second, fuller assessment is performed entirely online by educators and consists of five parts, designated A through E.

A: General suitability

B: Local suitability

C: User interface

Figure 1 Schematic showing the online assessment process where data handling is performed by CGI scripts.

D: Educational style

E: General review

Part A examines the general suitability of the resource and addresses the question of whether the material is correct, up to date, and of an appropriate level for a course known to the reviewer. If the material does not satisfy these criteria, the review can be stopped at this point and comments added if required.

Part B (Figure 2) provides information required for classification purposes and, where applicable, defines where the resource fits into the local curriculum.

The last three parts are standard questions on the user interface[23,24] (Table 2), the educational style[22] (Table 3), and a general review focused on hypertext aspects.[25] Answers are indicated on a four-point scale, as Yes/No responses, or as selections from a list of options. Additional text boxes are available to enter more information. A sample online review form is available on the Web.[26]

Stage 3: Supplementary Assessment

Resources that pass the Stage 2A assessment are evaluated for more general, non-educational quality factors such as copyright issues,[27] authorship,[28] stability, and maintainability.[29,30]

Pilot Implementation of the Methodology

The methodology was designed to be generic and applicable to the large range of subjects and teaching methods in medicine. For the pilot, the areas of interest were limited to clinical chemistry, radiology, and medical physics. This choice of topics partly reflected the interests of the participants but also provided examples of different types of medical teaching. In our institution, clinical chemistry is taught *en bloc* and provides a factual knowledge of pathophysiology, whereas the teaching of radiology is diffuse. Radiology is utilized in many areas of medicine and is much more concerned with interpretative and diagnostic skills. Medical physics is not taught to medical undergraduates but is delivered at a number of levels to practitioners in a range of professions. The subjects all cater to a wide range of students including medical students, other science graduates, and postgraduate students. Also covered are the different types of educational materials encountered on the Web: In clinical chemistry materials are mainly text-based with some graphics, whereas in radiology most materials are image-based, with the attendant challenges of image resolution, image quality, and the presentation of sequential and annotated images.
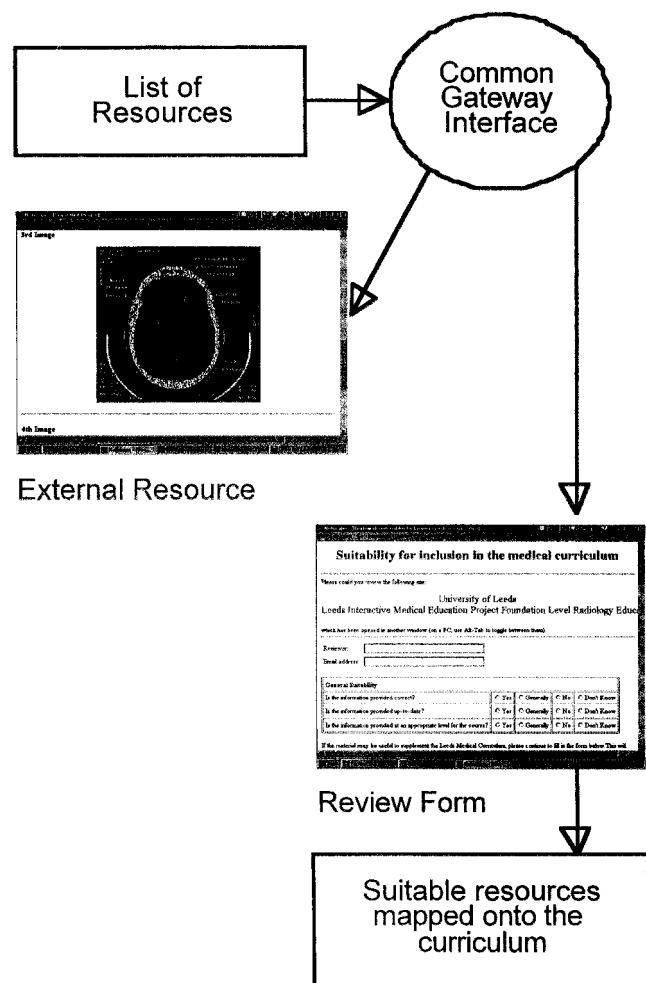
Figure 2 Screen shot of section of the online questionnaire covering assessment of resource and mapping to the local curriculum.

Table 2 ■

Questions Included in the User Interface Section of the Online Questionnaire

How compatible is the package for use from a cluster PC?
Was the package robust/reliable?
Is it possible to make teaching point with image quality/display available?
Please grade ease of navigation/operation.
Did you know where you were?
Did you know what you were expected to do next?
Please grade the speed of access.
Please grade the readability.

Table 3 ■

Questions Included in the Educational Style Section of the Online Questionnaire

How was the site organized? Hypertext or linear path?
Please grade the quantity of information.
Interactive graphics?
Type of material (textbook/tutorial/cases/lecture slides/simulation)?
Is this information better presented in a textbook?
Is this information better presented in a tutorial?
Is this information better presented in a lecture?
Is the information self-contained?
Was there any self-assessment?
If yes, was response mandatory or optional?
Method of question (MCQ/case study/short answers)?
Please grade question quality.
Method of response (online/mailed/none)?
Please grade feedback qualtiy.

Reviewers with domain expertise were recruited in each of the subject areas, and each used the assessment process on resources with the potential to support taught courses. Reviewers not closely involved with the project but familiar with using Web browsers needed only a page of instructions explaining the less familiar terms used on the review form. Those new to using the Web received a one-hour personal tutorial on both general Web matters and the assessment.

Results

The types of materials identified differed between the subject areas and are illustrated in Table 4. In clinical chemistry a total of 27 resources were identified, and 15 passed Stage 1 of the assessment. It was possible to use the online process on all of these, and this was applied by four reviewers. In spite of an encouraging number of resources identified, only two were found to be valuable locally following review. The main problems were a lack of depth to the material and a poor fit between the material and local course requirements. For example, although five reference books were identified, these took the form of operational laboratory handbooks. The majority of case studies held material on rare conditions graded for postgraduate use.

In radiology a total of 80 resources were identified, and 76 passed Stage 1 of the assessment. A number

of these contained more than one educational package. Much more material was found for experienced radiologists than for undergraduates. Five potential reviewers were recruited, but Stage 2 of the assessment was initially not possible because no local written curriculum for undergraduate radiology teaching was available during the study period.

In medical physics a total of 36 resources were identified, and 25 passed Stage 1 of the assessment. More resources were found in the newer imaging areas, such as magnetic resonance imaging, than in traditional topics such as physiologic measurement. Three reviewers were recruited, and ten sites were reviewed by two of these. On review, two sites were deemed suitable for immediate use; one site involved large files that required local mirroring to be worthwhile but did not require local support material; four would be useful only with additional support material, and of these two would need mirroring; and three sites were found to be incomplete and limited in scope and were probably never intended for external users. Each review took about 20 minutes—slightly more for the more complex sites where it took reviewers longer to familiarize themselves with the educational resources.

Discussion and Conclusion

Recent work[31] has shown that Web-based materials that are not integrated into a structured framework have limited educational value. The model presented here supports such a view, because the database of reviewed materials will allow material to be presented to users by a variety of interfaces, including one closely matching the taught components of the course. Simple cataloging of resources is insufficient in the teaching domain. It is only when potentially useful resources undergo full assessment that their true utility is clarified, especially with regard to local applicability. This is one of the main advantages of this methodology over simply providing users with access to a bank of search engines, since available search engines do not structure their catalogues of material in an educationally useful way.

The questionnaire asks about the fit to the local curriculum and thus best fitted topics where the teaching is highly structured and formal, as it is in clinical chemistry. However, most clinical chemistry material on the Web took the form of operational laboratory handbooks and specialist technical information. These offered little support for case-based teaching of the subject. The best materials were undoubtedly case-based, although these tended to focus on rare conditions that are of more relevance to postgraduate train-

Table 4 ∎

Types of Educational Material Identified in the Pilot Study Subject Areas

	Clinical Chemistry		Medical Physics		Radiology	
	%	No.	%	No.	%	No.
Annotated images	7	1	24	6	20	15
Case studies	33	5	0	0	0	0
Multiple-choice questions	7	1	0	0	0	0
Teaching files	0	0	0	0	53	40
Reference texts	33	5	44	11	20	15
Tutorials	20	3	32	8	8	6
Total no.		15		25		76

ees. The model was also appropriate to medical physics, but support material would be required to ensure the educational usefulness of the available resources. Again, mainstream materials were absent. In our institution, many clinical disciplines including radiology have unstructured undergraduate curricula. At the time of this study there was no list of course aims and objectives for radiology. This meant that for materials suitable for undergraduates it was hard to judge the relevance of resources to the curriculum in Stage 2B of the process, and our model was totally inappropriate.

Over the three subject areas, the results challenged some perceptions on the material available. Although we identified as many resources as expected, a lower proportion than hoped were judged to be of use locally. Potential users may also have to change their way of working in order to use Web-based materials —perhaps building up their own collection of resources designed to allow students to make the most of available online material. This is particularly true for interactive simulations,[32] where a user may need to be led through a number of examples to benefit from the resource. Reviewers found that exposure to the rigorous review criteria alerted them to educational and design factors, and as a result changes to working practice are already under way.

It was also notable that materials on the Web outside the field of radiology are not well-developed enough to allow for widespread use. Educators themselves are still in the learning phase and may mount materials experimentally rather than create a functional Web learning site. Mounting materials on the Web is also an endpoint for student projects, especially in medical physics. Thus materials are of highly variable quality, and the sites may be ill-maintained.

The reviews returned by assessors form the basis of a Web-based interactive database that can used by students and educators to locate resources by standard

fields and keywords. Educators are encouraged to develop their own Web-based introductory and support materials, and students may also be directed to resources from course- or lecture-specific pages or from an online teaching timetable. These local access methods would allow a centralized database to be used in more than one institution. Indeed, one way to approach the exponential increase in available materials would be to establish an international electronic forum for medical educators similar to the Computers in Teaching Initiative Centre for Medicine Guide,[33] where additions to the Web could be publicized, thus reducing duplication and helping the maintenance of databases of resources. Following this pilot work it will be necessary to evaluate the fuller implementation in terms of inter- and intra-reviewer reproducibility.

This review process does not evaluate the impact of learning itself, an area of some debate, as summarized by Anderson and Draper[15] in the area of computer-based learning. Although pre- and post-testing can show an effect, they do not help explain what caused the effect, why it occurred, and what learning process was affected by the new intervention.

Commercial Web users, newspapers in particular, are concerned about the copyright of materials made available on the Web. Resources such as LIME depend on the freedom to include hyperlinks to external resources, and this principle is currently under debate.[34] As the database grows we will ask permission from authors for their resources to be included. This approach will not only overcome potential copyright problems but will allow us to establish contact with authors with the potential for determination of stability and information on changes and additions.

In summary, we have established an online assessment methodology, which will generalize to other subject areas with a written curriculum and with minor modifications to other institutions. International differences in terminology and reference ranges limit the possibilities of global sharing of resources.

References ■

1. Friedman RB. Top 10 reasons the World Wide Web may fail to change medical education. Acad Med. 1996;71:979–81.
2. Chodorow S. Educators must take the electronic revolution seriously. Acad Med. 1996;71:221–6.
3. Stoker D, Cooke A. Evaluation of networked information sources. In: Information Superhighway: the role of Librarians, Information Scientists and Intermediaries. Proceedings of the 17th International Essen Symposium, October 24–27, 1994. Essen, Germany: Essen University, 1995:287–312.
4. Taubes G. Indexing the internet. Science. 1995;269:1354–6.
5. Ding W, Marchionini G. A comparative study of web search service performance. Proceedings of the ASIS Annual Meeting. Washington, D.C.: American Society for Information Science, 1996;33:136–42.
6. Norman F. Digital libraries: a quality concept. In: Arvanitis TN (ed). Proceedings of the European Congress of the Internet in Medicine, 1996, Brighton, England [CD-ROM]. Amsterdam, The Netherlands: Elsevier Science, 1997. Excerpta Medica International Congress Series no. 1138.
7. Health on the Net Foundation Web site. Available at: http://www.hon.ch/home.html. Accessed October 1997.
8. Friedman CP. The research we should be doing. Acad Med. 1994;69:455–7.
9. Computers in Teaching Initiative Centre for Medicine (CTICM) Update. 1997;8:32.
10. Van Ort S. Evaluating audio-visual and computer programs for classroom use. Nurs Educ. 1989;14:16–8.
11. Posel N. Guidelines for the evaluation of instructional software by hospital nursing departments. Comput Nurs. 1993;11:273–6.
12. Huber JT, Giuse NB. Educational software evaluation process. J Am Med Inform Assoc. 1995;2:295–6.
13. Clark RE. Dangers in the evaluation of instructional media. Acad Med. 1992;67:819–20.
14. de Dombal FT. Medical Informatics: The Essentials. Stoneham, Mass.: Butterworth-Heinemann, 1996.
15. Anderson A, Draper SW. An introduction to measuring and understanding the learning process. Comput Educ. 1991;17(1):1–11.
16. Glenn J. A consumer-oriented model for evaluating computer-assisted instructional materials for medical education. Acad Med. 1996;71:251–5.
17. Dailey DJ, Eno KR, Brinkley JF. Performance evaluation of a distance learning program. Proc 18th Annu Symp Comput Appl Med Care. 1994:76–80.
18. Harkin PJR. Quality assurance in CBL material. University of Leeds School of Medicine Web site. Available at: http://www.leeds.ac.uk/medicine/lime/curriculum/qacbl/qacbl.html. Accessed October 1997.
19. Computers in Teaching Initiative (CTI) Centre for Medicine, University of Bristol Web site. Available at: http://www.ilrt.bris.ac.uk/cticm/about.htm. Accessed October 1997.
20. Teaching & Learning Technology Programme Web site. Available at: http://www.tltp.ac.uk/tltp/index.html. Accessed October 1997.
21. Graham IS. HTML Sourcebook. New York: Wiley, 1996:68–74.
22. Laurillard D. Computers and the emancipation of students: giving control to the learner. Instr Sci. 1987;16:3–18.
23. Baecker R, Buxton WAS. Empirical evaluation of user interfaces. In: Readings in HCI: A Multidisciplinary Approach. San Mateo, Calif.: Morgan Kaufman, 1987:135–47.
24. Edmonds E. Human–computer interface evaluation: not user-friendliness but design for operation. Med Inform. 1990;15(3):253–60.
25. Knussen C, Tanner GR, Kibby MR. An approach to the evaluation of hypermedia. Compu Educ. 1991;17(1):13–24.
26. Sample online review form, University of Leeds School of Medicine Web site. Available at: http://www.leeds.ac.uk/

medicine/lime/docs/sample_review.html. Accessed October 1997.

27. Stern EJ, Westenberg L. Copyright law and academic radiology: rights of authors and copyright owners and reproduction of information. Am J Radiol. 1995;164:1083–8.

28. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet. JAMA. 1997;277(15):1244–5.

29. Ciolek TM. The six quests for the electronic grail: current approaches to information quality in WWW resources. Review Informatique et Statistique dans les Sciences Humaines (RISSH), Centre Informatique de Philosophie et Lettres. Liege, Belgium: Universite de Liege, 1996;1–4:45–71. Also available at: http://coombs.anu.edu.au/SpecialProj/QLTY/TMC/QuestMain.html.

30. Brandt DS. Evaluating information on the Internet. Comput Libraries. 1996;16:44–7.

31. Booth AG, Aiton JR, Bowser-Riley F, Maber JR. The BioNet project: beyond writing courseware. Biochem Soc Trans. 1996;24:298–301.

32. Woodward J, Carnine D, Gersten R. Teaching problem solving through computer simulations. Am Educ Res J. 1988; 25(1):72–86.

33. Computers in Teaching Initiative (CTI) Guide to computer-assisted learning packages. CTI Centre for Medicine, University of Bristol Web site. Available at: http://wwwdev.ets.bris.ac.uk/cticm/calguide.htm. Accessed October 1997.

34. Kleiner K. Surfing prohibited. New Scientist. 1997; 153(2066):29–31.