The Effectiveness of Hypermedia Instructional Modules for Radiology Residents

Steven G. Shaw

Roger Azevedo

Patrice M. Bret

AbstractThis case study details the development and subsequent field testing of training materials for residents in radiology. Eight modules were produced in a hypermedia format at the Montreal General Hospital by teams comprising medical faculty, residents and educational technologists. The contents of the modules address clinical, anatomical and radiological themes. The nature and functionality of the modules, and the process involved in their development, are described. The results of a six-month field test utilizing a post-test only, quasi-experimental design are also presented. The evaluation addressed three types of learning: perceptual discrimination and interpretation, factual knowledge and inferential or diagnostic reasoning. Overall, the results indicated no significant differences between hypermedia and traditional instruction on any of the three categories of learning and only a slight difference among three levels of residency. Implications for future developments are discussed.

Resume: Cette etude de cas porte sur le developpement d'un hypermedia educatif par 1'Hopital General de Montreal pour les residents en radiologie, ainsi que sur les tests d'apprentissage administres sur le terrain suite a 1'implantation du logiciel. Les huit modules, qui abordent des themes cliniques, anatomiques et radiologiques, ont ete produits par une equipe composee de medecins, de residents et de specialistes en technologic educative. La nature des modules, leurs fonctions et les Stapes de leur developpement sont decrites. Les resultats des tests, effectues sur une periode de six mois selon un design quasi-experimental avec *post-test,* sont presentes. Trois types d'apprentissage sont mesures : discrimination et interpretation perceptuelle, connaissances factuelles et inferentielles et raisonnement diagnostique. Les resultats n'indiquent aucune difference significative entre 1'enseignement traditionnel et le format hypermedia quant au trois categories d'apprentissage. II existe, toutefois une legere difference parmi les niveaux de residents. Les implications pour de futurs d^veloppements sont analysees.

During the period September 1991 to August 1992, eight hypermedia computer-based tutorials for radiology education were developed at the Montreal General Hospital. The design, development, production and evaluation of the modules was carried out by a team comprising medical faculty from McGill University, residents from McGHFs network of teaching hospitals, and faculty and students from the Graduate Programs in Educational Technology at Concordia University and Universite de Montreal.

Postproduction evaluation of the modules was conducted during the period spanning September 1992 to April 1993.

In this case study we explain the rationale for utilizing a hypermedia approach in this domain and the expected benefits. The context, including the residency program in radiology and existing traditional teaching methods, is outlined. We provide a brief review of related development and evaluation projects in radiology education and comment concerning their relationship to our own work. The particular approach to development which was employed is described briefly, including aspects concerning organization, processes, tools and funding. Observations concerning the strengths and weaknesses of the approach are provided. The characteristics and functionality of the modules are also described. We then outline the evaluation scheme that was followed, including both preproduction and postproduction formative evaluation phases, and provide details of the postproduction evaluation which took the form of a field test. Procedures, tools and analyses associated with the field test are presented. Results are discussed and recommendations for future research and development are elaborated. Overall, findings were positive, with residents who learned with hypermedia generally performing as well as those who experienced traditional lecture and demonstration methods.

Context

The project was carried out at the Montreal General Hospital under the direction of Patrice M. Bret, Chief of Radiology. The radiology program is a five year, highly selective specialization which follows accreditation as a general practitioner. There are only approximately two dozen residents at the five teaching hospitals in the Montreal area in this particular specialisation. The goal of the project was to develop and implement improved teaching methods in the radiology program.

Radiology candidates must develop knowledge and skills in four areas, namely: (a) perceptual discrimination and interpretation; (b) factual knowledge (e.g., incidence of different pathologies in different populations); (c) inferential or diagnostic reasoning, and; (d) patient and case management. All four forms of competence are developed and refined through supervised case loads. Traditionally, instruction for varieties one through three is also addressed through a lecture series and through "conferences". Conferences generally are one-hour sessions during which as many as five residents are required to diagnose individual cases before their peers and a staff radiologist who has prepared the cases.

This system of lectures and conferences and lectures is somewhat haphazard. Lectures, many of which are given by visiting staff, vary from year to year both in content and in quality of presentation. Conferences are important tools for learning, but only one or two residents are truly actively participating at each session. Our observations of conference sessions revealed a low level of attention from some non-participants. Also, preparation time for both conferences and lectures is quite substantial and staff radiologists have less time at their disposal for these tasks than is desirable.

Given the foregoing analysis, it was our belief that technology could play a role in improving the quality of instruction while also, in the long run, freeing up faculty members' time for case supervision and case work.

Hypermedia For Radiology Training

Hypermedia has enormous potential as an instructional tool in the domain of radiology (Ackerman, 1992; Greenes, 1992; Jaffe & Lynch, 1992; Lesgold and Katz, 1992), especially with regard to the first three objectives listed above: development of perceptual and inferential skills and acquisition of factual knowledge. More explicitly, computer-based hypermedia offers the following possibilities:

- design a highly interactive form of instruction that guarantees active learner participation;
- present images (CTs, MRIs, Doppler images, X-rays) and highlight them in different ways to train residents in perceptual discrimination tasks and to test interactively their abilities in these tasks;
- present cases and pose cases for solution;
- utilise flexible instructional strategies including simulation and case-based learning and, more generally, both discovery and tutorial modes;
- construct a seamless environment for study with interconnected on-line glossaries, bibliographies, abstracts and search functions;
- provide facilities for annotating, saving, and printing images; generating bibliographies; saving and printing comments and notes, and;
- exploit thematic connections among concepts and cases through the structural flexibility of hypermedia systems, a capability which Feltovich et al, (1989); Spiro et al, (1988); and Spiro et al, (1991) argue is especially important in ill-structured domains.

By exploiting these possibilities, the following benefits were expected to accrue from the project:

- improved curricula in radiology;
- standardization of the quality and content of core curricula;
- improved learning through sound instructional design and the use of interactive learning strategies;
- on-going improvement of instruction through formative evaluation (including evaluation based on on-line collection of data);
- increased understanding of variables affecting learning in hypermedia in medical education;

dissemination of tools (a shell) and the transfer of development methodologies to radiology programs throughout Canada;
reduced instructional preparation time for faculty, and;
increased availability of explicit training to interns, with a growing repository of instructional and reference materials available through flexible scheduling.

Review Of Related Projects

Despite the apparent potential of hypermedia for the domain of radiology, there are relatively few projects which have been evaluated and reported. Those projects which have been reported are limited in scope and there are weaknesses apparent in the methodologies that were employed in their evaluation.

Among the studies which have been conducted we find the following: Moore, Kathol, Zollo and Albanese (1993) assessed the effectiveness of a videodisc film file compared with a conventional film file for teaching radiology to medical students. One hundred and thirty-four fourth-year medical students studied 116 cases selected from the American College of Radiology Learning File. Material comprised 58 skeletal cases and 58 chest cases. One-half the students studied the skeletal cases on film, the other half on videodisc. The conditions were then reversed for the chest cases. There was no significant difference in learning between the two conditions for either set of cases. However, students utilizing film reported their perception of a superior learning experience in terms of amount learned, convenience of use and ability to detect lesions. This is an interesting study insofar as it provides some justification for the use of lower resolution media for teaching purposes. However, the interactive and advanced navigational features afforded by the technology were not exploited in the design of this particular project.

D'Allesandro, Galvin, Erkonen, Albanese, Michaelsen, Huntley, McBurney and Easley (1993) compared the instructional effectiveness of a hypermedia textbook (HyperLung) concerning a lung disease with that of a lecture. HyperLung is a sophisticated hypermedia "pop-up" textbook comprising a table of contents, discrete chapters, indices, a dictionary, on-line testing and capabilities for text search and annotation. Forty-nine staff and residents in radiology were randomly assigned to receive instruction either through HyperLung or via a lecture. Both groups received the same content and each was tested before and after instruction. There was no significant difference in learning between the two groups.

The evaluation conducted by d'Allesandro *et al* is perhaps the one closest to the evaluation conducted at the Montreal General hospital which is reported in this article. However, the latter has a smaller, presumably more homogeneous sample (residents only), and also provides longer exposure to the treatment condition.

Jaffe, Lynch and Smeulders (1989) developed a hypermedia program on echocardiography that was intended to be used as the primary instructional tool for achieving an intermediate level of clinical expertise. It consists of a user-controlled learning environment with random access to 54,000 images and 1,200 clinical items. This program has proved successful in providing a uniform basic curriculum in echocardiography. After five to ten hours of independent study with this resource, residents are said to "have achieved an intermediate level of expertise and need less tutoring from the attending physician" than previously (Jaffe *etal,* 1989, p. 479). An objective qualitative assessment of skill acquisition was reported to be in progress. However, the authors do not reveal how the improvements in skill and performance mentioned above were measured.

Wenzel and Gotfredsen (1987) studied retention of theoretical knowledge after computer-assisted instruction in intraoral radiography. No significant difference was found between treatment and control groups on immediate or delayed (three months, 18 months) post-test measures, though both groups showed significant differences on pretest versus immediate post-test measures. Instrumentation comprised 20-item multiple choice tests.

Starkshall, Riggs and Lowther (1986) report the development and informal assessment of a computer-aided instructional module for radiological physics. The program appears to be a prototype. Informal feedback collected from physicists, residents and technology students focused on educational value and user friendliness. The feedback reported was positive and supports the use of the technology in the curriculum.

Several features distinguish the work conducted at the Montreal General Hospital from the projects and evaluations documented above. In particular, our project involves several learning objectives, a variety of materials and themes presented through one standard interface, and a relatively long duration with some 20 hours of material delivered over a six month period and four and one-half hours of formal testing.

While there are weaknesses in the design of the evaluation and in the instrumentation, as discussed later in this article, overall the approach is more rigorous than the evaluations described above. Arguably, our evaluation provides more reliable and more detailed evidence of the applicability of hypermedia to instruction for radiology programs than do the other evaluations which have been reported in the literature. Important features of the evaluation plan include an analysis of the impact on learning by type of objective, residency level and instructional method, a field implementation, and a collaborative learning situation for the field implementation, which is described below.

The Tutorials

The modules developed during the project cover the following topics: (a) Cystic Tumours of the Pancreas, (b) Ectopic Pregnancy, (c) Physics of X-ray Filters, (d) Tumours of the Posterior Fossa in Children, (e) Anatomy of the Temporal Bone, (f) Diagnosis of Arthritis, (g) Arterial Stenosis and Occlusion Viewed Through Ultrasound, and (h) Echogenic Liver Nodules. Each module contains between two and four hours of content, divided into chapters. A standard interface and a standard set of functions were designed for use in conjunction with all the modules and these were incorporated in a shell programmed in Supercard on the Macintosh, which served as the basic authoring tool. A medical market scanner (an Omniscanner equipped with a back light) was used for digitizing directly radiological images. Sound (for Doppler) was digitized using MacRecorder. Standard illustration programs such as MacPaint and Adobe Photoshop were used for creating graphics. Macromedia's Director program was used for animations. The digital video standard was Quicklime.

The shell was developed by a staff radiologist and the interface was subjected first to expert review and then subsequently, with some content in place, to testing with the assistance of graduate students from Concordia and Universite de Montreal to determine usability. Three graduate students from Concordia were employed in the role of "end-users" during the usability testing which was carried out with a representative module. Two usability testing strategies were employed: task-based evaluation and an evaluation grid. In the task-based component of the exercise, students were required to obtain certain specific information from the module or to perform a particular task (e.g., paste an image to the scrapbook and print it). Data was collected through observation and on-line trace routines to identify problems with the tools and interface presented to the user. Using the evaluation grid, students evaluated the module along several dimensions using an instrument comprised of Likert-scale items measuring ease of use, quality of graphics and the degree of usefulness of specific features and options.

A staff radiologist and one senior resident were then assigned to develop the content and finally actually to produce each module. During production, the staff radiologist who developed the shell was available to provide technical support to the different development teams. Each module underwent periodic formative evaluation for content accuracy and for consistency in presentation style and graphics. Consistency within and among modules was verified at regularly scheduled group presentations of work in progress attended by staff developers and instructional technologists.

The core content of each module corresponds to a specific lecture scheduled during the 1992-1993 sessions. The staff radiologist assigned to develop each module was the same individual who would deliver the lecture.

Some modules cover all three aspects of radiology training referred to above. Some cover only two (leaving out diagnosis or perceptual training), while one module, the one concerning physics of X-rays, covers only factual or theoretical knowledge.

Modules incorporated digitized images (of computed-tomography images [CT], X-rays, magnetic resonance images [MRI] and ultrasound), animations, digitized sound (e.g., doppler effects), text, and graphics (medical illustrations). Figures 1 through 3 show typical screens from the Arterial Stenosis and Anatomy of the Temporal Bone modules, incorporating text, images and sound. Words in bold are hot-linked to both glossaries and references. Images are enhanced with animation and highlights illustrating the text. A standard menu bar provides access to navigational tools (chapters, pages, page-forward, page-back, end-of-section, beginning-of-section, quit and a trace of the navigational path), a context-sensitive help facility, and a set of utilities labelled "goodies".

Figure 1
Text and digitized image from the Arterial Stenosis module.

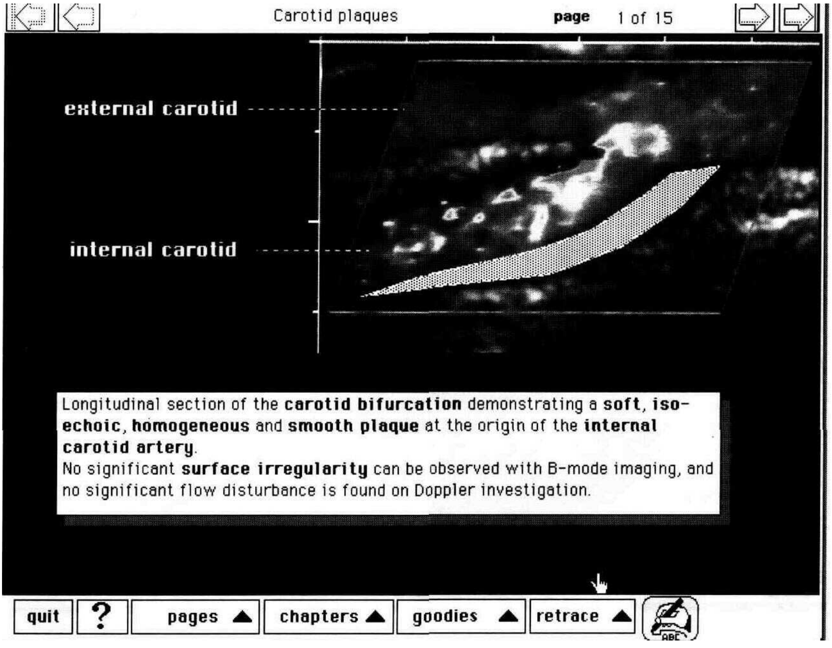


Figure 2
Screen exhibiting use of animation and digitized sound.

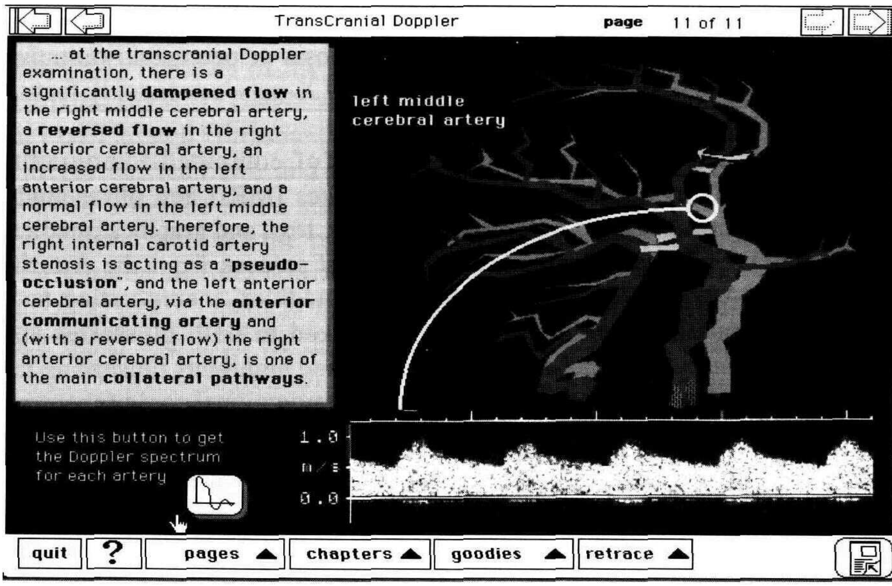
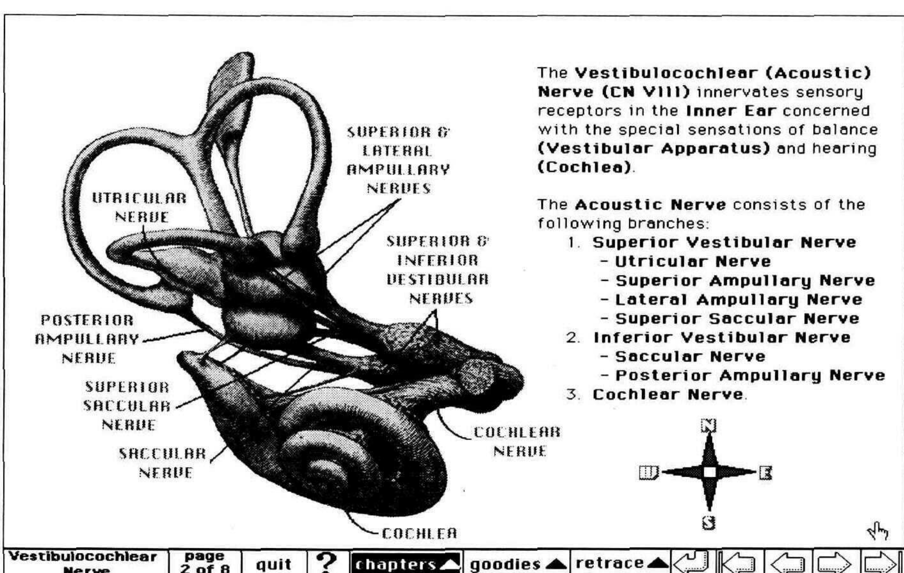


Figure 3
Typical use of medical illustration overlayed with "hot-spots".



Users can locate information and navigate via three alternative search functions. First, all modules incorporate key-word indexing and search. Two models were employed for key-word search facilities. Figure 4 illustrates a simple key-word search function, while Figure 5 shows a key-word search facility with cross-links to the glossary and bibliography. In the larger modules glossaries, indices and on-line bibliographies are cross-referenced and they are all available from a single screen displaying multiple windows (see Figure 6). All modules also include a second indexing device for navigation which is referred to as a "navigation board" (see Figure 7). The navigation board is accessible on any page from the "goodies" pop-up menu. It provides the user with a list of key words and pages related to the topic currently displayed. By clicking on a page or chapter appearing in the list presented in the navigation board the user may jump directly to that particular location in the module. Finally, some modules also incorporate full-text search as an alternative strategy for locating specific information.

Figure 4
Simple key-word search function.

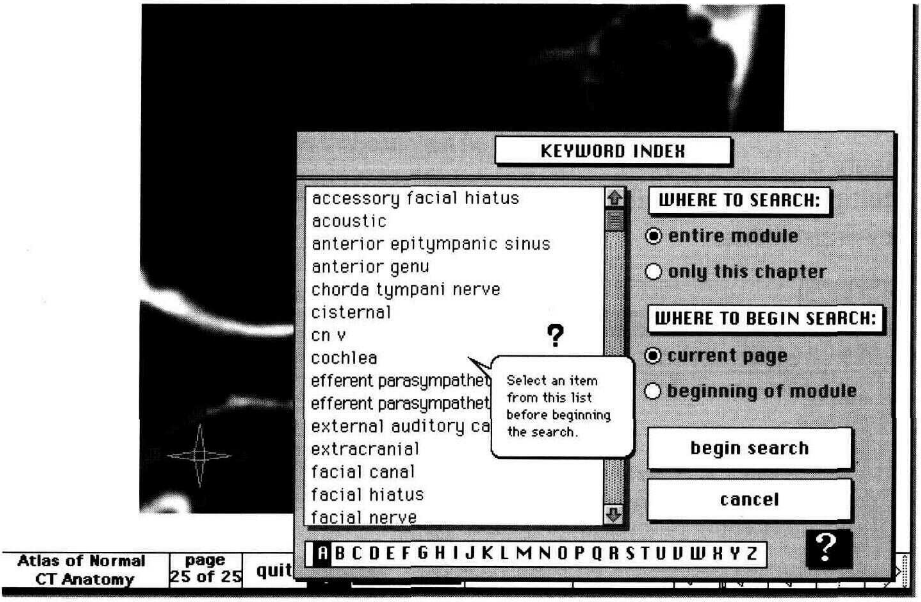


Figure 5
Key-word search facility with cross-links with glossary and bibliography.

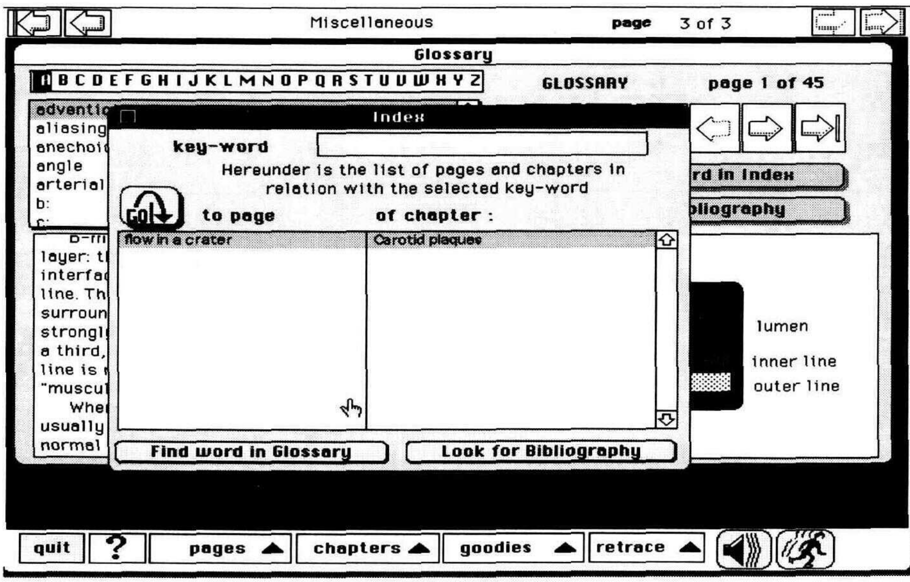


Figure 6
Multiple windows exhibiting connections among bibliography, glossary and key-word index.

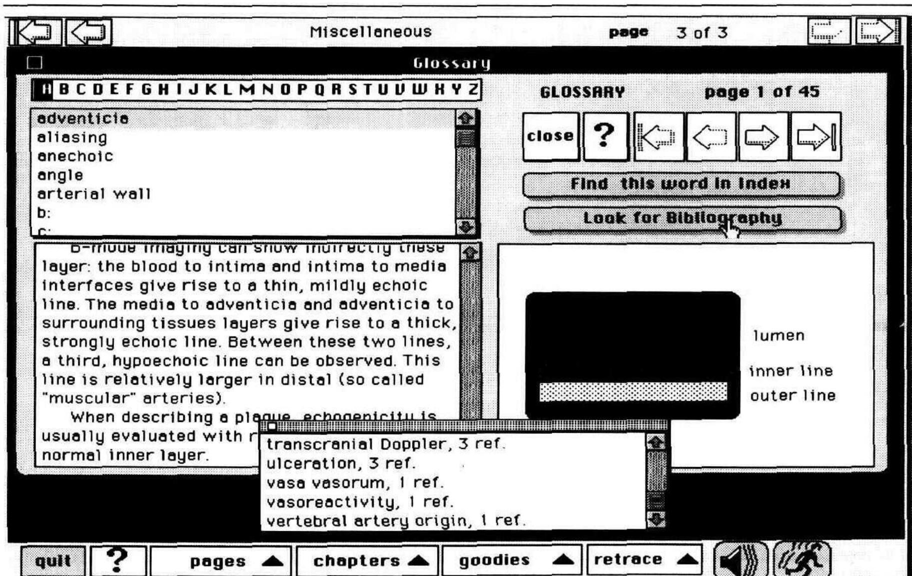
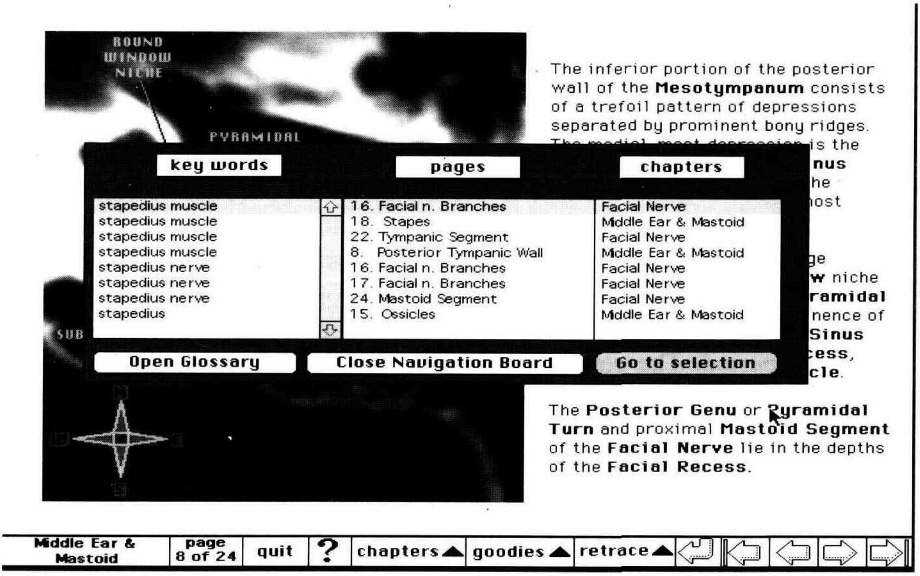
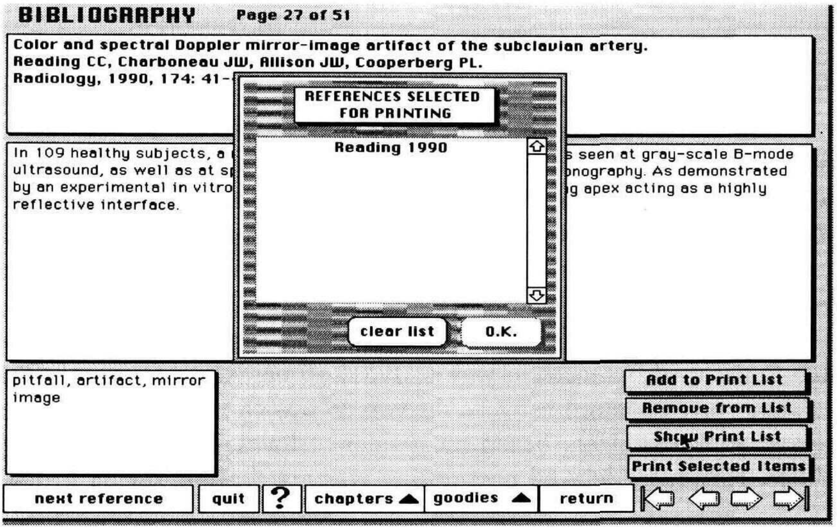


Figure 7
Navigation board.



Bibliographies include on-line abstracts and there are provisions for assembling and editing lists of references that can then be printed (see Figure 8)

Figure 8
Bibliography with on-line abstracts.



Other standard features include: a scrap-book where users can file images and text; a notebook, graphic tools and a text editor to annotate contents pasted to the scrap-book; print functions for scrap-book contents and notes; self-evaluation exercises; electronic bookmarks, and; dribble files with time stamps.

Although the modules are relatively sophisticated from a functional or technical standpoint, they were developed and produced largely by medical staff. Instructional technologists provided minimal production support while contributing to a greater extent to the design. Final production values were reasonably professional as judged by the instructional technologists involved.

The project was kept on schedule by the Chief of Staff of Radiology. This is a significant point in the organization of the project. It is estimated that over 6,000 person hours were spent in development and production phases over a three-year period. The cooperation and motivation of staff radiologists and residents, who are heavily burdened with cases and teaching responsibilities, could only be secured and maintained over time through the priority assigned to the project by the Chief of Staff.

In the following sections we report the study conducted to evaluate six of the eight modules and conclude with recommendations for future research and development activities.

Method

Subjects

Twenty-four radiology residents at McGill's five Montreal area teaching hospitals participated. There were 12 females and 12 males. The sample represents all residents available during the period of the study. Participation in the study was urged by the Chief of Staff of Radiology, and no resident declined to participate.

Design

A field test involving six of the eight modules was conducted over a six month period. A stratified random sampling divided the available group of residents (n=24) into two groups, with equal numbers of senior and junior residents in each. One group was assigned to the hypermedia condition, the other was to experience traditional lectures. No pretesting was employed; prior knowledge was treated as a random variable.

Originally the plan was to use a 2 x 2 x 3 x 5 x 8 mixed repeated measures design to assess differences in various types of learning (perceptual, factual, and diagnostic reasoning) between the two groups on immediate and delayed post-test administration, based on residency level, for each of the eight modules. However, delayed post-testing was not feasible owing to logistic problems associated with assembling residents from the various teaching hospitals at the same time on so many occasions, and sample size and variable mortality experienced over the six post-tests did not permit a true repeated measures analysis of the data. Therefore, six independent t-tests were conducted to compare groups on post-test scores and six one-way ANOVAs were conducted to analyze differences based on residency level. Independent t-tests were employed to investigate which individual items differentiated the two treatment groups.

Instruments

Post-tests were designed to be as similar in format as possible. Each post-test comprised multiple choice and short answer items. One test (Arterial Stenosis and Occlusion Viewed Through Ultrasound) included photographic prints of ultrasound images with specific areas that were to be identified and outlined by residents. Each test was constructed by a lecturer with assistance from the instructional technologists. Tests were reviewed to ensure adequate sampling of the content, and to verify that content was contained in both lectures and hypermedia. The hypermedia modules contained additional content not intended to be presented in the lectures. This content was excluded from the post-tests.

Pilot testing of several of the post-tests was conducted with residents who were not included in the study in order to evaluate clarity and duration. Guidelines for format and length of the post-tests were created on the basis of the validation of the first post-test. There was not a sufficient number of residents available to allow for formal pilot testing of tests for all six modules.

Scoring of test results was carried out by the lecturer responsible for the content. Tests were identified only by a code number in order to ensure impartial evaluation. For non-objective test items (e.g., short answer diagnoses or precise identification of pathological features in images) with a range of possible scores, the lecturer provided an evaluation grid, and items were scored by both the lecturer and an instructional technologist. Inter-rater reliability for this scoring procedure was set at .90 and this level was achieved.

Procedure

On the six testing days, which occurred at one month intervals, all residents were assembled at the Montreal General Hospital in a meeting room. They were instructed to go either to the lecture hall or the hypermedia lab, according to their random assignment to conditions. The lectures were scheduled to last one-hour. The hypermedia sessions were concurrent with the lectures with the same duration. At the end of the one-hour presentation period, all subjects were reunited in the meeting room and the tests were administered during a 45 minute period.

Subjects were monitored by the instructional technologists during the administration of the tests. They were instructed to place the last four digits of their phone number on the cover sheet for identification purposes and to record their residency level (year one through five) on the back of the last page. The completed tests were returned to the invigilators, who recorded the completion time.

On the first testing day, the objectives of the study were explained to participants. In particular, they were informed that the results of the testing were not a formal component of their evaluation in the residency program, and that the primary objective was to evaluate the two instructional systems. Those not in the hypermedia condition were informed that access to the hypermedia materials would be provided to them after testing.

A survey revealed that no participant in the hypermedia condition had ever experienced computer-based instruction and also that there was little experience of application software other than clinical applications. Hence, in preparation for the hypermedia condition participants were provided with a twenty-minute training session. For this session, a self-instructional module was prepared using the same shell and interface as encountered in the medical education modules. During the training session the instructional technologists were present to provide additional support to the participants.

In the hypermedia condition, participants worked in groups of two or three at each workstation. They communicated freely during the sessions, within their groups. On-line traces recorded what content was accessed and the time spent in the module. During the hypermedia sessions participants were free to access any content, but they were informed which chapters corresponded to the lecture material and would therefore be directly relevant to the subsequent test.

The order of the modules was determined by the availability of the lecturers. There was no obvious relationship of dependency among the contents of the different modules.

Results

While eight modules were developed, only six were included in the study, owing to time constraints and availability of team members and subjects. The six were: (a) Cystic Tumors of the Pancreas, (b) Ectopic Pregnancy, (c) Physics of X-ray Filters, (d) Tumors of the Posterior Fossa in Children, (e) Anatomy of the Temporal Bone, and (f) Arterial Stenosis and Occlusion Viewed Through Ultrasound.

Descriptive statistics (sample sizes, group means, standard deviations, maximum scores on post-tests) for the six modules are presented for the two conditions in Table 1 . The SPSSx program *Frequencies* was employed to assess assumptions of univariate normality and to detect outliers. Tests for univariate normality were satisfactory.

Table 1
Descriptive Statistics for the Six Hypermedia Modules.

	Sample Size		Mean (SD)		
Hypermedia	Hypermedia Group	Traditional Teaching Group	Hypermedia Group	Traditional Teaching Group	Maximum Possible Score on the Posttest
1) Artereal Stenosis and Occlusion Viewed Through Ultrasound	10	14	31.60 (10.46)	32.29 (8.18)	64
2) Physics of X-Ray Filters	11	13	33.18 (8.68)	31.62 (7.08)	50
3) Ectopic Pregnancy	10	12	26.30 (4.64)	29.83 (3.76)	52
4) Cystic Tumors of the Pancreas	6	15	15.83 (3.82)	17.20 (2.51)	42
5) Anatomy of the Temporal Bone	8	13	14.25 (3.88)	15.08 (5.28)	30
6) Tumors of Posterior Fossa in Children	6	13	6.50 (1.23)	9.08 (1.32)	10

Of the six independent t-tests comparing post-test scores (see Table 2), only the one associated with Tumours of the Posterior Fossa was significant $t[17] = 4.04$, $p < .05$. However, irregularities regarding the length, format (short answer, only) and design of the post-test (redundant items), lead us to discount this result.

Table 2
Independent t-tests Comparing Groups on Post-test Scores.

	Mean (SD)	T-Test	
Hypermedia Module	Hypermedia Group	Traditional Teaching Group	Value
1) Arterial Stenosis and Occlusion Viewed Through Ultrasound	31.60 (10.46)	32.29(8.18)	0.18
2) Physics of X-Ray Filters	33.18(8.68)	31.62(7.08)	0.49
3) Ectopic Pregnancy	26.30 (4.64)	29.83 (3.76)	1.97
4)Cystic Tumors of the Pancreas	15.83(3.82)	17.20(2.51)	0.97
5) Anatomy of the Temporal Bone	14.25 (3.88)	15.08(5.28)	0.38
6) Tumors of the Posterior Fossa in Children	6.50(1.23)	9.08(1.32)	4.04*
* p < .05			

Of the six one-way ANOVAs (see Table 3) conducted to investigate differences by level of residency only the one addressing Physics of X-ray Filters was significant $F(3,19) = 3.4$, $p<.05$. Subsequent post hoc analyses in the form of independent t-tests indicated the difference related to item five, a factual recall item which asked the resident to identify the material used for the node of an X-ray unit, the reasons why the material is used and at what kVp characteristic X-rays emerge. Junior residents outperformed seniors on this item.

Table 3
One-way ANOVAs Comparing Residency Level on Post-test Scores.

Hypermedia Module	F value
1) Arterial Stenosis and Occlusion Viewed Through Ultrasound	$F(4,19)=2.54$
2) Physics of X-Ray Filters	$F(3, 19)=3.44$*
3) Ectopic Pregnancy	$F(3, 18)=0.34$
4) Cystic Tumors of the Pancreas	$F(4, 16)=1.56$
5) Anatomy of the Temporal Bone	$F(3, 12)=0.18$
6) Tumors of the Posterior Fossa in Children	$F(3, 13)=0.71$

$p<.05$

Independent t-tests administered for items in each module to determine which individual items discriminated between the groups revealed seven additional significant items (see Table 4).

Table 4
Item analysis.

Hypermedia Module	Posttest Question Number	Mean Score of Hypermedia Group	Mean Score of Traditional Teaching Group	T-test Value
1) Arterial Stenosis and Occlusion Viewed Through Ultrasound	Ql	3.60	0.79	5.12*
	Q14	0.70	2.82	.361*
	Q15	0.20	4.43	4.05*
	Q16	0.70	1.25	0.20*
2) Physics of X-Ray Filters	Q5	4.00	3.07	2.21*
3) Ectopic Pregnancy	Q3	1.30	2.25	2.82*
	Q7	0.60	1.00	2.45*
	Q15	1.20	0.50	2.97*

* $p < .05$

Discussion

Basically, we found no significant differences between instructional conditions or by residency level. As indicated above, only the module entitled Tumours of the Posterior Fossa revealed differences based on instructional condition and this result should be discounted because of problems associated with the length, format and design of the test. Only one module yielded significant differences based on levels of residency: Physics of X-rays. However, this module did not cover all three types of learning objectives. Furthermore, the single significant difference between residency levels should be interpreted in the context of multiple tests of significance drawn from essentially the same sample, with the concomitant increased likelihood of a spurious finding of a difference.

As acknowledged earlier, there was variable mortality so samples are not strictly identical across the tests and a repeated measures analysis was consequently contra-indicated. The mortality is not overly large, however, and we assume that the absences are random. Unfortunately, it is not possible to guarantee that there will be no mortality in an evaluation of this kind, given the many commitments of residents, unforseen hospital emergencies or illness among residents themselves. Variable mortality is the price that must often be paid to gain the external validity that comes with a field study conducted over an extended period of time.

The results do not appear generally to be due to a ceiling or floor effect. Most modules have mean scores between 50 and 65 percent of total possible scores, with reasonably large standard deviations (see Table 1). The lowest mean scores are found in Cystic Tumours of the Pancreas where they are about 40 percent which might be interpreted as a floor effect. This module did not have any items which discriminated between treatment conditions.

An important aspect of the design of the evaluation is duration. One would expect any initial positive or negative effect on performance due to the novelty of the treatment condition to dissipate over time. In fact, the actual results belie any such effect, since the only significant difference between treatment and control conditions did not occur until residents were tested with the fourth module, well into the field study, and since this finding has also been discounted.

Table 4 lists those post-test items for which the two treatment groups differed significantly. The items concerned fall into two categories: recall of factual information and visual discrimination tasks. Two factual items favoured the hypermedia treatment, while two favoured the control group. In the module concerning Arterial Stenosis and Occlusion, the control group outperformed the hypermedia group on three visual discrimination tasks. In the

Physics of X-ray Filters module, the hypermedia group outperformed the control group on a theoretical question.

While it is difficult to draw conclusions from these few results, it is interesting to note three things. First, the results for the three visual discrimination tasks are interesting. The hypermedia system provides more images than were presented during the lecture, and these images are highlighted and annotated in sophisticated ways. We therefore expected the hypermedia group to outperform the lecture group on this class of tasks. The results suggest that the impact on transfer of the difference in resolution between images displayed in the software and the regular film output should be investigated. In the visual discrimination tasks presented during testing, residents worked from copies of film. Film resolution is over 1000 dots-per-inch (dpi), while images incorporated in the modules are scanned down to 600 dpi and then displayed on a 72 dpi screen. This is a reduction in quality that may have a significant impact on feature identification and discrimination. The transfer issue is also interesting given that the technology is changing in the field and increasingly radiologists are being required to work from digital images displayed on a CT screen, rather than from film, regardless of how they received their training.

Secondly, the hypermedia module addressing the physics of X-ray filters made extensive use of dynamic visual displays (animations and Quicklime clips) to illustrate theoretical principles. The superior performance of the hypermedia group on a factual item addressing a theoretical issue is thus encouraging, insofar as it indicates the effectiveness of this strategy.

Thirdly, the general non-significance of the results provides a solid justification, from the standpoint of instructional effectiveness, for the use of the particular hypermedia format that was investigated. The long history of non-significant findings in instructional media or methods comparison studies would tend to support the prediction of a non-significant result in our field study. However, the amount of content available in the hypermedia condition is much greater than in the lecture, and on-line data indicates a considerable amount of material extraneous to the lectures was accessed by residents in the hypermedia condition. The lectures are highly structured and focused in comparison with the hypermedia material and they essentially teach to the tests that were employed. Thus, there was some reason to expect that there might be significant differences found in favour of the lecture condition.

The overall results of the evaluation are consistent with other studies which have investigated computer-based applications in radiology education. The methodology and sample are different than other reported studies. Other studies have included larger but less homogenous samples, different technologies (e.g., videodisc), lack details concerning instructional design approach or granularity, are merely one shot post-test designs, or use different measures (or do not report measures).

Despite the limitations in our evaluation, as per the preceding discussion, the present field study provides strong evidence that technology-based instruction can be as effective as traditional lectures in delivering complex or refined subject matter to radiology residents.

Recommendations For Future Research And Development

With a no significant difference result, one can conduct a simple cost-effectiveness or cost-benefit analysis to decide whether the technological approach is worthwhile. While the initial effort to develop the eight modules was considerable, the benefits are obvious, as outlined above. The material is available on demand, while the content and quality of delivery is standardized. With the present shell it is a relatively easy matter to add new chapters or content to existing modules and, with the development and production expertise acquired within the radiology department during the course of the project, creation of new modules should be possible with less effort. Cost-effectiveness is enhanced by the circumstance materials can be shared with radiology programs across Canada. On the other hand, it must be acknowledged that the "freeware" approach limits the possibilities of developing and marketing the materials in a commercial form, to generate revenues for further work.

In assessing the strategy for funding this project, it should be remembered that a major goal was to develop a standard shell with a fixed interface which might achieve widespread use. One of the major problems associated with the growing body of published commercial software for medical education is that there is a great deal of variation in functionality and interface design. Standards and conventions yield efficiencies for both developers and end-users (learners). We reasoned that a robust design which could "satisfice" requirements for delivering content related to radiology education would be a useful contribution to software development in this area.

Funding for this project was minimal. Forty-thousand dollars was provided by the Canadian Radiological Association and the Sterling Winthrop pharmaceutical company. Despite this limited funding, over 20 hours worth of material was developed. The project demonstrated the feasibility of in-house development and production within a teaching hospital, with limited start-up capital. For the moment, until the market matures, development of usable products for smaller specialized market segments will continue to depend on the efforts of committed in-house developers.

Future evaluations should utilize truly standardized post-test instruments that have been assessed for reliability. This was not strictly possible in our evaluation given the limited number of residents available within the McGill teaching hospitals. A more representative sampling of the three types of learning objectives would also be desirable.

As mentioned previously, the study involved group utilization of hypermedia. No attempt was made to exploit this circumstance in the design of the modules and their functions, and groups were formed randomly and changed to some extent over time, based on absenteeism. Further studies might control for this variable, or investigate the impact of the variable on performance and attitudes.

The decision to utilize groups in the delivery of the hypermedia materials was based on three factors: (1) the nature of radiological activities during residency, which are typically group-based, (2) differences among participants regarding experience with computers, and (3) availability of appropriate computer resources (we didn't have 15 high-end Macintosh computers).

In a strict sense, the group dimension is clearly a confounding element in the field study (did residents learn from the hypermedia materials or from their interactions with peers?). But from an evaluation standpoint, the issue is not critical - unless one wishes to generalize to use of the materials on a strictly individual basis.

Finally, future developments of the applications should entertain the following recommendations:

The programs should exploit hypertext capabilities to a greater extent, incorporating more mini-cases (following Spiro et al, 1988; 1991).

Enhancements to program functionality should be based on a detailed analysis of realistic learning and clinical activities in radiology, following principles of cognitive apprenticeship and situated learning. Such an analysis would allow implementation of a cognitive apprenticeship strategy with variable levels of support for learning and performance (Collins, Brown & Newman, 1989) and would be in line with recent thinking about the need to accommodate situated aspects of complex problem-solving activities in ill-structured domains, including social and cultural dimensions (Clancey, 1993).

Further recommendations from formative evaluation of the interface, concerning aspects such as standardization of location of visual versus textual information, should be implemented.

The possibilities offered by multiple windows for promoting visual acuity based on multiple image comparison should be explored. The present software offers only limited comparison of images with two per screen, or with comparison images on successive screens or interchangeable within a window.

The issue of the impact of lower resolution grey-scale images on visual perception tasks and its impact on transfer to operational radiological conditions should be investigated.

References

Ackerman, M. J. (1992). The videodisc in medical education. In J. C. Honeyman & E. V. Staab (Eds.), *Syllabus: A special course in computers for clinical practice and education radiology* (pp. 169-170). Chicago, IL: Radiological Society of North America Publications.

Collins, A., Brown , J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. In L. Resnick (Ed.), *Knowing, learning and instruction: Essays in honour of Robert Glaser* (pp. 453-494). Hillsdale, NJ: Lawrence Earlbaum Associates.

Clancey, WJ. (1993). Guidon-manage revisited: A socio-technical systems approach. *Journal of Artificial Intelligence in Education, 4(1),* 5-34.

D'Allesandro, M. P., Galvin, J. R., Erkonen, W. E., Albanese, M. A., Michaelsen, V. E., Huntley, J. S., McBurney, R. M., & Easley, G. (1993). The instructional effectiveness of a radiology hypermedia textbook (HyperLung) versus a standard lecture. *Investigative Radiology, 28(1),* 643-648.

Feltpvitch, P. J.,. Spiro, R.J., & Coulson, R.L. (1989). The nature of conceptual understanding inBiomedicine: The deep structure of complex ideas and the development of misconceptions. In D. A. Evans & V. L. Patel (eds.), *Cognitive Science in Medicine,* 112-172. Cambridge, Massachusetts: The MIT Press.

Greenes, R. A. (1992). Computer-based education and decision support in radiology. In J. C. Honeyman & E. V. Staab (Eds.), *Syllabus: A special course in computers for clinical practice and education radiology* (pp. 177-183). Chicago, IL: Radiological Society of North America Publications.

Jaffe, C. C., & Lynch, P. J. (1992). Computer-aided instruction for radiological education. In J. C. Honeyman & E. V. Staab (Eds.), *Syllabus: A special course in computers for clinical practice and education radiology* (pp. 171-176). Oak Chicago, IL: Radiological Society of North America Publications.

Jaffe, C. C., Lynch, P. J., & Smeulders, A. W. M. (1989). Hypermedia techniques for diagnostic imaging instruction: Videodisc echocardiography encyclopedia. *Radiology, 171,* 475-480.

Lesgold, A., & Katz, S. (1992). Models of cognition and educational technologies: Implications for medical training. In D. A. Evans & V. L. Patel (Eds.), *Advanced models of cognition for medical training and practice* (pp. 171-190). Berlin, Germany: Springer-Verlag.

Moore, T. E., Kathol, M. H., Zollo, S. A., & Albanese, M. A. (1993). Comparison of a videodisc system with a conventional film file for medical student teaching. *Investigative Radiology, 28(10),* 969-973.

Spiro, R. J., Feltovich, P. J., Jacobson, M. J., & Coulson, R. L. (1991). Cognitive flexibility theory, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains. *Educational Technology, 31(5),* 24-33.

Spiro, R. J., Coulson, R. L., Feltovich, & P. J., Anderson, D. K. (1988). *Cognitive flexibility theory: Advanced knowledge acquisition in ill-structured domains.* Paper presented at the 10th annual conference of the Cognitive Science Society, Montreal, Quebec, Canada.

Starkshall, Riggs, & Lowther. (1986). A Computer-aided instructional module for radiological physics. *International Journal of Radiation Oncology, Biology, Physics, 12,* 421-426.

Wenzel & Gotfredsen. (1987). CAI in intraoral radiography. *Journal of Dental Education,* 57(5), 244-245.

AUTHORS

Dr. Steven G. Shaw is an Associate Professor in the Educational Technology Program, Department of Education, Concordia University.

Roger Azevedo is a doctoral student in the Department of Educational & Counselling Psychology, Faculty of Education, McGill University.

Dr. Patrice M. Bret is the Radiologist in Chief at Montreal General Hospital and Professor of Medicine at McGill University.