

Linear Versus Web-Style Layout of Computer Tutorials for Medical Student Learning of Radiograph Interpretation¹

Martin V. Pusic, MD, MA, Vicki R. LeBlanc, PhD, Steven Z. Miller, MD

Rationale and Objective. We sought to determine which is more effective in increasing skill in radiograph interpretation: a linear (PowerPoint-style) computer tutorial that locks the student into a fixed path through the material or a branched (Web-style) version that allows random access.

Materials and Methods. We prepared a computer tutorial for learning how to interpret cervical spine radiographs. The tutorial has 66 screens including radiographs or graphics on almost every page and five unknown radiographs for the student to interpret. One version (linear) presents the material in a linear sequence with the unknown radiographs heading up “chapters” detailing an important aspect of the task. In the second (branched) version, the same 66 screens were accessed through hyperlinks in a frame beside the unknown radiographs. One hundred thirty-nine medical students at two sites participated in a randomized single-blinded controlled experiment. They interpreted cervical spine images as a pretest and then completed one of the two tutorial versions. Afterward, they did the same examination as a post-test.

Results. The tutorial was successful, in both layouts, in improving the subjects’ ability to interpret cervical spine radiograph images (effect size 2.1; 95% confidence interval 1.7–2.5). However, the layout did not make a difference to their gain in ability. Students in the linear group completed the tutorial in 17% less time ($P < .001$) but were slightly less likely to rate the tutorial as “valuable.”

Conclusion. For these novice learners, computer tutorial layout does not affect knowledge gain. Students may be more satisfied with the linear layout, but in time-pressured situations, the Web-style layout may be preferable because it is more efficient.

Key Words. Computer Assisted Instruction; Educational Research; Hypermedia; Instructional Design; Learner Controlled Instruction; Radiology; Randomized Controlled Trial; Undergraduate Medical Education.

© AUR, 2007

Acad Radiol 2007; 14:877–889

¹ From the Department of Pediatrics (M.V.P.), University of British Columbia, Vancouver, BC, Canada, and the Centre for Research in Education (V.R.L.), University of Toronto, Toronto, Canada; and the Department of Pediatrics, College of Physicians and Surgeons, Columbia University, New York, NY (S.Z.M.). M.V.P. is currently at the Division of Pediatric Emergency Medicine, Columbia University, Department of Pediatrics, PH1-137, 622 West 168th Street, New York, NY 10032. This study was funded by a Canadian Institute of Health Research and Association of Canadian Medical Colleges Research in Medical Education grant. Received Mar 28, 2007; Revision received April 16, 2007; accepted April 18, 2007. **Address correspondence to** M.V.P. e-mail: mvp19@columbia.edu

© AUR, 2007

doi:10.1016/j.acra.2007.04.013

The skill of radiograph interpretation includes recognition of visual features and the use of a procedure to ensure that all relevant information is considered. Students learn this skill through a combination of instruction and practice. Computer tutorials are an excellent medium for this learning since they facilitate aggregation of representative radiograph examples, and they allow novel instructional designs that are more interactive than textbooks or radiology teaching files (1). There are a large number of online resources and CD-ROM computer tutorials available for this particular skill (2). Unfortunately, rigorous evaluations of their educational impact are lacking.

These computer tutorials are generally laid out in one of two formats. The first type is a linear PowerPoint style

of presentation that is *behaviorist* in that the goals are prespecified and the path of the student is largely determined ahead of time by the instructor. The *behaviorist* philosophy specifies that learning best takes place when a desired change in behavior is prespecified (3–5). The educator provides a program of instruction specific to named objectives. Typical examples of instructional designs based on this “top-down” philosophy are lectures and programmed instruction. This method has the advantage of being very efficient.

The second type of layout commonly used for computer-based learning materials is a branched nonlinear one typified by most Web sites. Instead of locking the student into a linear sequence, the instructional designer sets up an interface that allows students to pick and choose the content they wish to access and the order in which they do so. This design can be more *constructivist* in nature when it allows the student to manipulate the materials at will to actively construct his or her knowledge from the materials provided (6). The emphasis is placed on having the learners discover new knowledge on their own instead of having it presented to them. Examples of designs based on this “bottom-up” philosophy are interactive learning environments (such as computer simulations) and problem-based learning (6).

The purpose of our study was to determine the effect of presenting an example topic via presentation layouts that are based on these two different philosophies. We held constant the content and basic instructional strategies while varying the overall method of presentation. The students either received programmed instruction via a PowerPoint-style linear presentation, or we presented the material in an interactive learning environment style where the student solved unknown problems with the content available as hyperlinked Web-style reference material.

Using cervical spine radiograph interpretation as an example, our null hypothesis was that if we held content constant, a Web-style (branched) instructional design would not differ from a more behaviorist linear layout in aiding medical students to develop the skill of radiograph interpretation.

METHODS

Participants

Participants were final-year medical students on their pediatrics rotation recruited at one of two sites: the British Columbia Children’s and Women’s Health Centre and the Morgan Stanley Children’s Hospital of New York.

We presented the study to them during their orientation to their rotation. The study material is relevant to their learning goals for the rotation. Following their participation in the study, we debriefed them, gave feedback, and gave them a book gift certificate worth \$20. The institutional review boards of both sites approved the study.

Computer Tutorials

We created a computer tutorial entitled “Interpretation of Pediatric Cervical Spine X-rays” using SumTotal Toolbook Instructor (Mountain View, CA). This software is an authoring language specifically developed for creating computer tutorials. Initially, we created the tutorial to accomplish prespecified learning objectives. Specifically, for patients who required cervical spine radiographs (CSXRs) after a traumatic event, we wanted students to be able to use specific visual features to distinguish which radiographs are abnormal. This first tutorial presented the material in a linear fashion like a PowerPoint lecture. We then created a second branched version of the tutorial that centered on identifying abnormalities on unknown images. These contrasting layouts are shown diagrammatically in Figure 1.

Linear tutorial

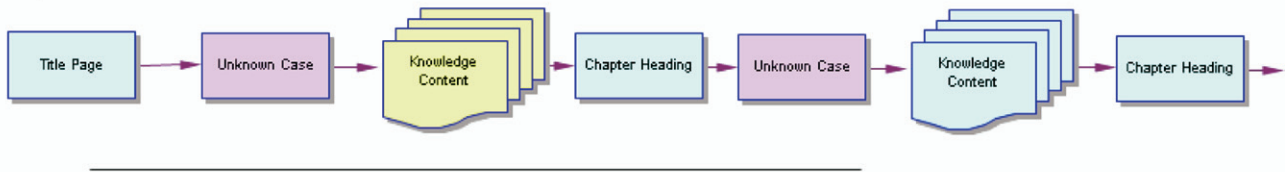
This tutorial is made up of 66 consecutively presented screens of text, schematics, anatomical drawings, and radiographs. The material is divided into six chapters: a review of the anatomy and then chapters (alignment, bones, count, dens, soft tissues) describing the steps required to complete the procedure of interpreting a CSXR. Within the first few screens of each chapter, the tutorial presents the students with an unmarked radiograph and asks them to locate the abnormality by pointing with the mouse cursor (termed *Unknown Cases*). An example of an unknown case is shown in Figure 2. Feedback is given based on the student’s response, and then the tutorial presents expository material reinforcing the learning concept. While the students can go backward and forward, they cannot significantly alter the order of presentation of the material.

Web-style tutorial

In the second tutorial, we used the same 66 screens but presented them differently (Fig. 3). The tutorial presents the six unknown cases serially while the remaining expository screens are available for access, in any order, behind onscreen links in a vertical frame occupying the lefthand side of the screen. The links group the expository screens

Linear Layout

All pages presented in sequence



Web Layout

Unknown case screens presented in sequence

Knowledge content screens available at all times behind links

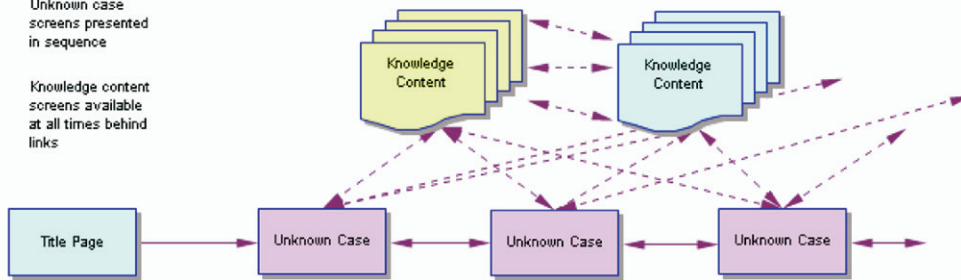


Figure 1. Schematic representation of the layouts of the two tutorials. In the linear group, the content of the tutorial was presented using serial screens of information. Unknown images are presented in sequence as one of the first screens for a given section of the tutorial. For example, the "Alignment" section begins with an image for which the key feature was malalignment of the cervical spine. For the Web-layout group, the content of the tutorial was presented using serial unknown images. The tutorial included the same number of unknown images but they were grouped together to form the backbone of the tutorial. Knowledge content could be accessed in any order from any of the unknown images. For each of the unknowns, a hyperlink was included that lead the student to the appropriate section for learning about the given anomaly.

by headings similar to the chapter headings of the linear tutorial. In addition, the students are able search the expository screens by key word and have them displayed in a separate window. The feedback screen for each unknown image has a hyperlink to the appropriate section of the content.

Experimental Design

The design of this study was a single-blinded randomized block comparison of the two tutorials using pre- and post-testing. We randomly assigned students to one of the two tutorials. The main outcome measure was the improvement in their ability to correctly classify 10 CSXRs. Secondary outcome measures included their learning efficiency (change score divided by time spent with the tutorial) and a measure of satisfaction. We determined the number of screens accessed by the students as well as their path through the tutorials.

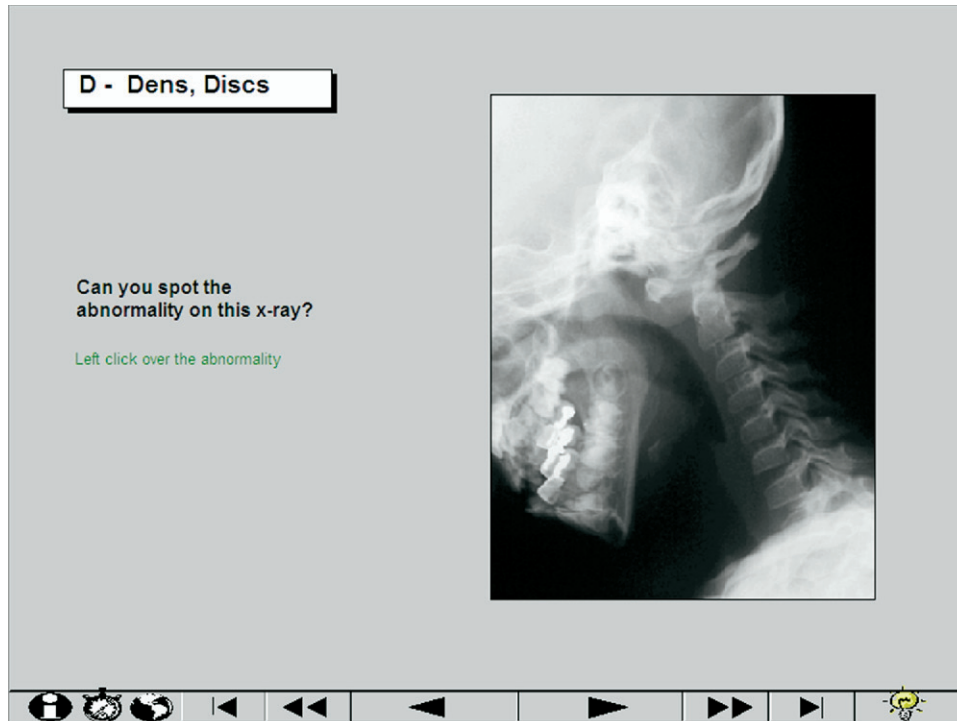
Outcome Measure

To develop the examination, we searched the medical records of the Children's and Women's Health Centre of British Columbia for all patients with a discharge diagno-

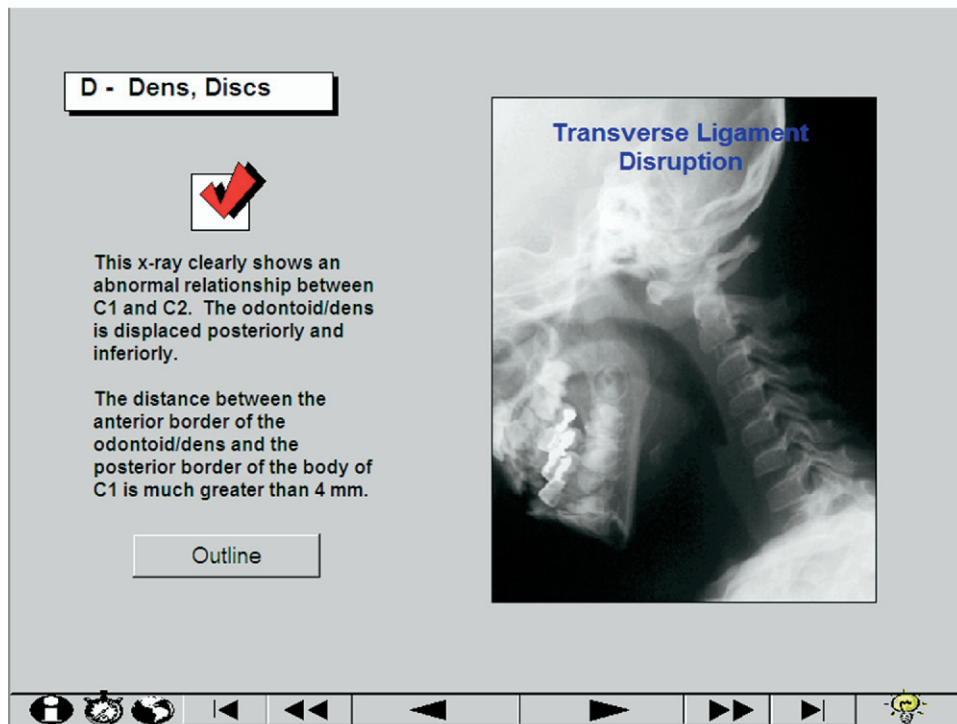
sis ICD-9 coded for cervical spine fracture between 1985 and 2001 (140 cases). Reviewing their charts, we pulled the initial radiographs at presentation for those patients who had a traumatic injury (70 cases, 138 images). We digitized the radiographs and retained those where the radiograph was nonobvious but showed a preoperative bony abnormality documented on the treating radiologist's report (35 images). From these images, we chose three or more examples, based on their representativeness and clarity, of each of the five major abnormalities addressed in the tutorials for a candidate set of 17 images.

We showed the 17 images to 5 medical students, 10 residents, and five pediatric emergency medicine fellows and recorded the trainees' responses. We calculated item difficulties for each image. Two of us chose the final 7 abnormal images based on which images (a) had progressively increasing scores with increasing trainee level and (b) showed the best discrimination index. We retained a minimum of one image for each of the five major abnormalities.

The objective of a radiologist in examining images from injured patients is to classify the patients into one of two categories: abnormal (potentially injured bone or liga-

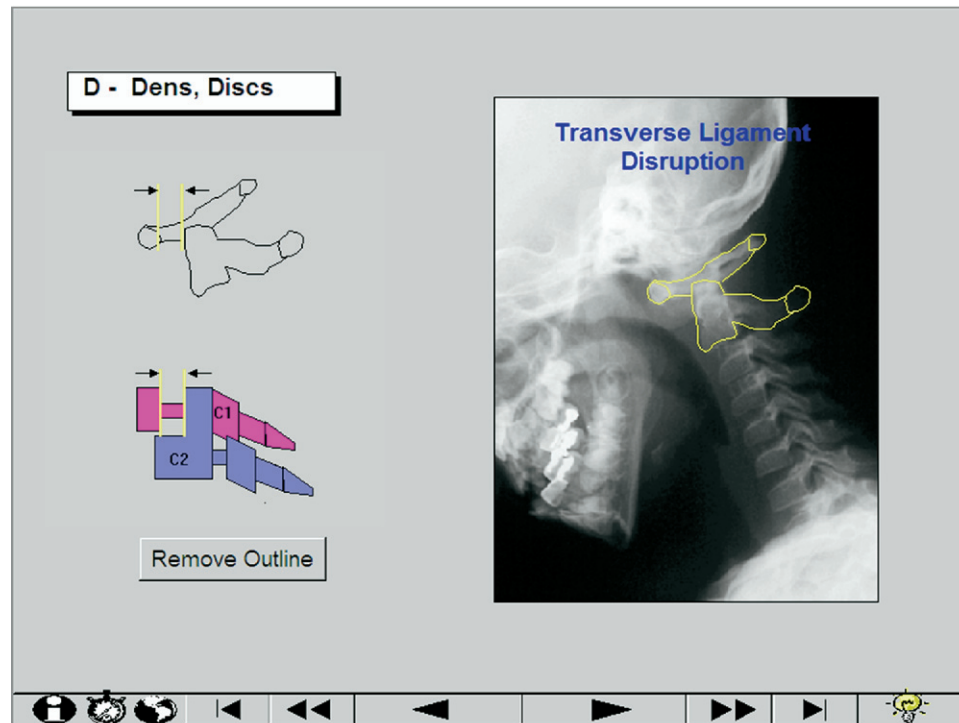


a.



b.

Figure 2. Three screen captures from the linear cervical spine x-ray computer tutorial. (a) First, an unknown case is presented from the linear tutorial. The student is asked to click with the left mouse button over a feature that suggests that this radiograph may be abnormal. (b) Second, after the student clicks on the unknown screen, the tutorial first presents a textual explanation of the features. (Fig 2 continues)



c.

Figure 2 (continued). (c) Third, clicking on the “Outline” button of the textual explanation reveals this visual explanation where the features on the radiograph are segmented and interpreted.

ments of the spine) or normal. In practice, more than 97% of inspected images are normal (7). To make our subjects’ task more representative of the real-world situation, we mixed in three normal images from children in our practice. We had the final 10 computer images reread by an independent pediatric radiologist blind to the initial radiologist’s interpretation. These independent interpretations were all concordant with the initial report.

The study subjects’ task was to specify whether a given image was normal or not; if abnormal, they had to specify the exact nature of the abnormality. One can argue that being able to specify that an image is abnormal, without specifying the exact level of the injury, has value because the patient is kept immobilized until a definitive judgment is made by either neuroradiologists or neurosurgeons. We pre-specified that we would examine the results using two possible scores for the main outcome variable: (a) score using correct distinction of normal from abnormal and (b) score using correct distinction of normal from abnormal plus specification of the level of the lesion for the seven abnormal images. Our main outcome would be the improvement in test score, using the method of scoring that yielded the most reliable test instrument after item analysis.

Procedure

The study procedure was carried out in a 1-hour session for each medical student group on rotation in the pediatric emergency department at the two institutions. After completing informed consent, the students completed the entire procedure at a single computer workstation. They completed the pretest and then, based on a random number function, the computer program seamlessly randomized the student to one of the two tutorials, thus ensuring allocation concealment. The students completed each phase of the procedure without time constraints. Immediately after completing the tutorial, they reinterpreted the same CSXRs and rated the following statement on a Likert-type scale: “I found this tutorial to be a valuable learning experience.”

Data Collection

Research assistants collected the data at the two sites. The students completed the tests and the tutorials in continuously proctored computer labs at individual workstations. The students’ interpretations of the images were coded by a research assistant blind to study group assignment. Two attending physicians resolved any disagree-

Figure 3. Screen captures from the Web-style layout computer tutorial. All screens can be accessed through hyperlinks arranged in a frame along the left-hand side of the screen. The right-arrow button takes the student to the next unknown case. The radiograph shows the same case of C1C2 malalignment as seen in Figure 2.

ments. The students' paths through the tutorials were tracked using a custom program that generates time-stamped entries to a log file each time the student accesses a new screen. The students' attitudes towards the tutorial was assessed using a "pop-up" question in between two of the questions of the post-test.

Data Analysis

Item scoring

The test was made up of 10 images: 3 normal and 7 abnormal. Students were assigned a 1 for each of the three normal images they correctly classified as normal and a 0 if they incorrectly classified it abnormal. For the 7 abnormal images, each image had only one abnormal

visual feature. The students were assigned a 1 for each item where the feature was correctly identified and a 0 if it was incorrectly identified.

Item analysis

We evaluated the items in our examination using classic theory item analysis. For each of the 10 items, we determined the point-biserial correlation (PBC) corrected so that it represents the correlation of the item with the total score with that item removed, as recommended for tests with fewer than 25 items (8). We did not prespecify minimum cutoff values for PBCs but instead proposed to use them to identify items that could be dropped to maximize total test reliability as measured by Cronbach's α (see later).

Reliability analysis

To determine the reliability of our test, we determined the overall Cronbach's α (unstandardized) of the test using the "stacked" pre- and post-tests according to the method of Wright et al. (9). To create the most reliable instrument possible, we dropped items that had poor PBCs. To determine overall reliability, we calculated the reliability of the subjects' difference scores (post-test score minus pretest score). We used a well-recognized formula based on the individual reliabilities of the examination and the variances of the students' scores (10). The reliability of difference scores is known to be less than the corresponding single-test reliability coefficient (10).

Outcome analysis

The test scores were submitted to a mixed-design analysis of variance, with the tutorial assignment as a between-subject variable and the time of test (pre-post) as a repeated-measure. The secondary measures of interest (number of screens accessed, path followed through the tutorials, and the students' attitudes) are described with the relevant descriptive statistics. Exploratory analyses were conducted to determine whether any of these variables contributed to the degree of changes in accuracy from the pretest to the post-test. Statistical significance of the student satisfaction responses was determined using the χ^2 distribution, while comparisons of time spent on the tutorials were done with the Student t -test. Number of screens accessed is not normally distributed, so comparisons were made using the Mann-Whitney U test. All analyses were done before unblinding the group assignment. The study is powered to detect a 10% absolute difference in the post-test score between the two groups assuming $\alpha = .05$, $\beta = .95$ and a standard deviation for both groups of 15%. This would require 57 students in each group. We could reasonably expect our within-groups design to result in a more powerful study design than a between-groups design. This made the likelihood of a Type II error small. We calculated the effect size for our measures using Cohen's d (11). The 95% confidence intervals are based on the noncentral t -distribution and calculated using the ESCI software (12).

Student path through tutorial

Each time a student accessed a new page in either tutorial, a line was written to a computer log file listing the time the page was accessed and the title of the page. To determine whether the subjects' paths through the two different tutorials differed, we wrote a Microsoft C# com-

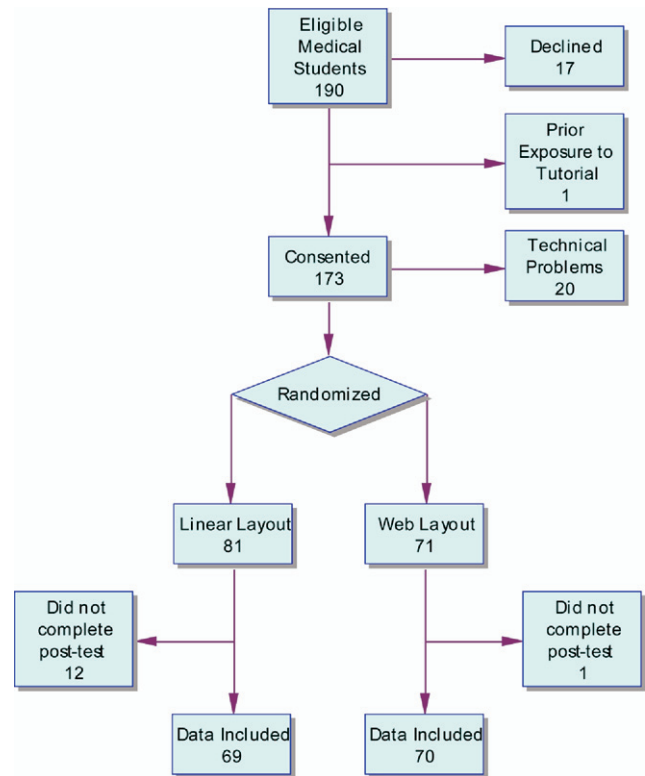


Figure 4. Study flow diagram.

puter program that read in each student's log file and converted it into a one-dimensional grid where each cell represents a single page viewed by the subject. Pages from a given chapter were represented as a single color. For example, pages showing "Unknown" images were black; "Anatomy," red; and so on. This gave us a visual representation of each subject's path through the tutorial, which could then be inspected for qualitative differences.

RESULTS

We carried out the study over a 10-month period beginning September 2002. Over 10 sessions, we enrolled 139 students, 51 at Columbia University and 88 at University of British Columbia. The study flow diagram is shown in Figure 4. We approached 190 students in total. Seventeen declined to participate. One was excluded because he had previously seen the computer tutorial. Twenty could not complete any part of the program due to a software problem during one session. Thirteen could not finish the protocol during another session when the

Table 1
Measured Outcomes for the Students Comparing the Two Intervention Groups

	Branched Group (n = 69)	Linear Group (n = 70)	Overall (n = 139)	Statistical Significance
Scores: precisely identifies lesion (marks of seven abnormal findings)				
Pretest (SD)	0.9 (1.0)	0.8 (0.9)	0.9 (1.0)	
Post-test (SD)	3.9 (1.2)	3.9 (1.2)	3.9 (1.2)	
Difference (SD)	3.0 (1.4)	3.1 (1.5)	3.1 (1.4)	
Effect size*: pretest versus post-test (95% CI)	2.1 (1.7 to 2.6)	2.1 (1.7 to 2.5)	2.1 (1.7 to 2.5)	F (1,137) = 615; P < 0.001
Effect size*†: branched versus linear (95% CI)	-0.06 (-0.39 to +0.27)			F (1,137) = 0.3; P = 0.61
Time on tutorial (min), mean (SD)	24.3 (9.7)	29.2 (8.2)	26.8 (8.8)	95% CI diff (2.3 to 8.0)
Screens viewed, mean (SD)	112 (61)	135 (33)	124 (50)	95% CI diff (7 to 40)
Time per screen viewed (s), mean (SD)	14 (5)	13 (4)	14 (4)	P = NS
Learning efficiency: Marks of 10 improved from pretest to post-test per minute spent on tutorial (SD)	0.12 (0.13)	0.06 (0.07)	0.09 (0.11)	P = 0.004‡

Scores indicates whether the student was able to indicate the exact lesion reported by the radiologist.

*Effect size is expressed as Cohen's *d*.

†Positive effect size favors branched version of tutorial.

‡Mann-Whitney *U* test used as learning efficiency variable not normally distributed.

computer lab was evacuated for a fire alarm. As a result, 139 completed the entire procedure, 69 in the Web-layout group.

Instrument

Item analysis

We calculated the item difficulties and point biserial correlations (PBCs) for all 10 items under the preintervention and postintervention conditions. The item difficulties ranged from +0.07 to +0.57 and were higher (less difficult) in the post-test condition.

We used the PBCs to determine which items to keep for an optimal examination. We noted that there was a considerable difference between images that show an abnormality and distracter normal images. The PBCs for the abnormal images were acceptable ranging from +0.10 to +0.49. They remained approximately the same for both the pre- and post-test. Whereas the seven abnormal images all had positive PBCs, all three normal images had negative PBCs, indicating that subjects who scored highly on the whole examination scored poorly on these items.

Reliability of instrument

The stacked reliability of the 10-item examination was poor at 0.21. Dropping the three items with poor PBCs (i.e., the three normal images) resulted in a considerable

improvement of the reliability to 0.71. Unless otherwise specified, all remaining analyses are based on the seven-item instrument. The reliability of the difference scores using the seven-item instrument was 0.21.

Outcomes

Classifying the images by specific lesion

For the seven-item test, including only the abnormal images, the students significantly improved their pre- to post-test scores from a mean of 0.9 image to almost 4 images read correctly out of seven (overall effect size for computer tutorial 2.1; 95% CI 1.7–2.5) (Table 1).

Our intervention (the layout) did *not* differentially affect the increase in knowledge from pre- to post-test (effect size for layout: 0.0; 95% CI -0.39 to +0.27). Improvement for the linear-layout group was 0.8 to 3.9 (of 7), whereas it was 0.9 to 3.9 for the Web-layout group (see Table 1).

The tutorial was most effective for teaching the concepts of C1C2 malalignment on the Odontoid view, the measurement of the predental space, and the necessity of fully visualizing all seven cervical vertebrae (see Table 2). As discussed in the item analysis section, the reliabilities of the three normal images were poor. Student performance on these images actually degraded after the tutorial going from an average score per item of 0.63 down to 0.31 (effect size -0.64).

Table 2
Examination Scores Before and After the Computer Tutorial, Based on the Radiograph Finding

Finding (# of questions)	View	PreTest Score Mean (SEM)	PostTest Score Mean (SEM)	Difference (95% CI)	Effect Size for Computer Tutorial
Normal (3)	Lateral	0.63 (0.02)	0.31 (0.02)	-0.32 (-0.26, -38)	-0.64
C1/C2 Malalignment (1)	Odontoid	0.13 (0.03)	0.77 (0.04)	+0.64 (0.55, 0.75)	1.3
Increased peduncular space (2)	Lateral	0.29 (0.03)	0.83 (0.02)	+0.54 (0.47, 0.60)	1.1
Vertebral Subluxation (2)	Lateral	0.02 (0.01)	0.27 (0.03)	+0.25 (0.19, 0.30)	0.70
Vertebral body crush (1)	Lateral	0.03 (0.01)	0.14 (0.03)	+0.11 (0.05, 0.17)	0.39
Inadequate film—C7 not visible (1)	Lateral	0.08 (0.02)	0.81 (0.03)	+0.72 (0.64, 0.80)	1.44

Scores represent the proportion of subjects who were able to correctly identify the level and nature of the abnormality.

Classifying the images as normal versus abnormal

When the students only had to specify whether an image was normal or not, their scores were higher but all of the statistical comparisons yielded results of similar magnitude and direction.

Time and learning efficiency

Students randomized to the linear tutorial spent more time on the tutorial (29.5 versus 24.3 minutes; 95% CI diff 2.3–8.0). The Linear group did considerably more screens than did the Branched group (135 vs. 112; 95% CI diff 7–40). The mean time per screen was normally distributed and did not differ between the groups, nor was it correlated with knowledge gain. As a result, the *learning efficiency* [(post-test score minus pretest score)/time on tutorial] of the Branched group was greater (0.12 versus 0.06 mark/min; $P = .004$).

Subjective rating

As sometimes happens in other walks of life, the less efficient group (linear) seemed happier. They were more likely to rate the tutorial 5/5 on a scale rating their agreement with the statement "I found this tutorial to be a valuable learning experience" (81% versus 55%; $P = .002$). Only 5 of the 139 students rated the tutorial anything less than a 4 (4 = agree, 5 = strongly agree).

Qualitative analysis of paths

Figure 5 is a visual representation of the paths taken by the students through the two tutorials. As expected, the students took substantially different paths through the material depending on the layout. The linear group each did the tutorial in a similar fashion, gradually working through each chapter. In the branched group, there was considerably more heterogeneity in the paths taken. For example, some of the students went through all of the

unknown images first and then went through the content, whereas others (fewer than 30%) essentially recreated the linear tutorial by sequentially going through the content screens.

DISCUSSION

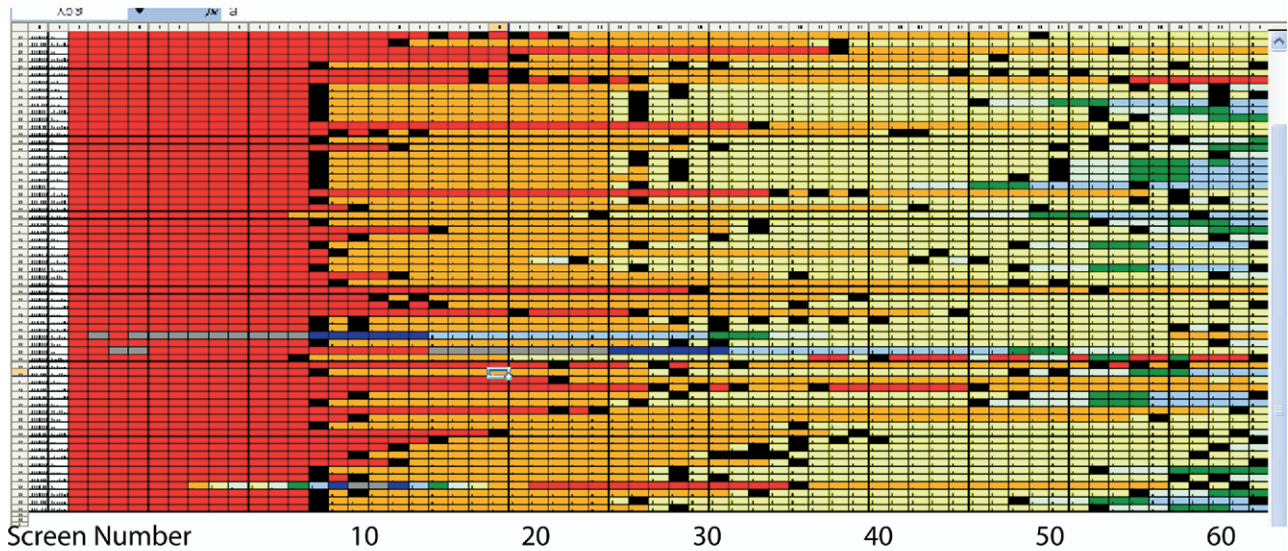
There are a large number of computer-based learning materials available for learning radiology interpretation for both undergraduate and postgraduate medical education (13, 14). Excellent examples include CD-ROMS from the American College of Radiology (15) as well as online materials at several sites (14), including Aunt Minnie (16), ELERA (17), and COMPARE (18). In a survey with a 24% response rate, Angle et al. (19) report that more than 85% of radiology residency programs have some sort of digital teaching file. There have been recent calls to increase the use of PACS systems to generate locally customized teaching databases of digital images (20–22).

There are a number of studies of radiology computer-based learning materials (CBLMs). Several radiological trainee satisfaction surveys have shown that trainees are generally positively disposed to online radiology learning (23, 24) although not universally so (25). Several media-comparative studies have compared CBLMs with other instructional media. D'Alessandro and colleagues (1, 26) showed as early as 1993 that a radiology computer tutorial could be as efficacious as a lecture, and others have replicated their findings in other radiology education contexts (27–31). In general, these studies have shown that computer tutorials can be at least as efficacious as lectures and even small-group sessions in raising test scores. Most have shown good subject satisfaction with CBLMs.

LEGEND

Colour	Content Chapter
y	Anatomy
a	Alignment
b	Bones
c	Count
d	Dens Discs
s	Soft Tissue
u	Unknown Film

LINEAR LAYOUT



WEB-STYLE LAYOUT

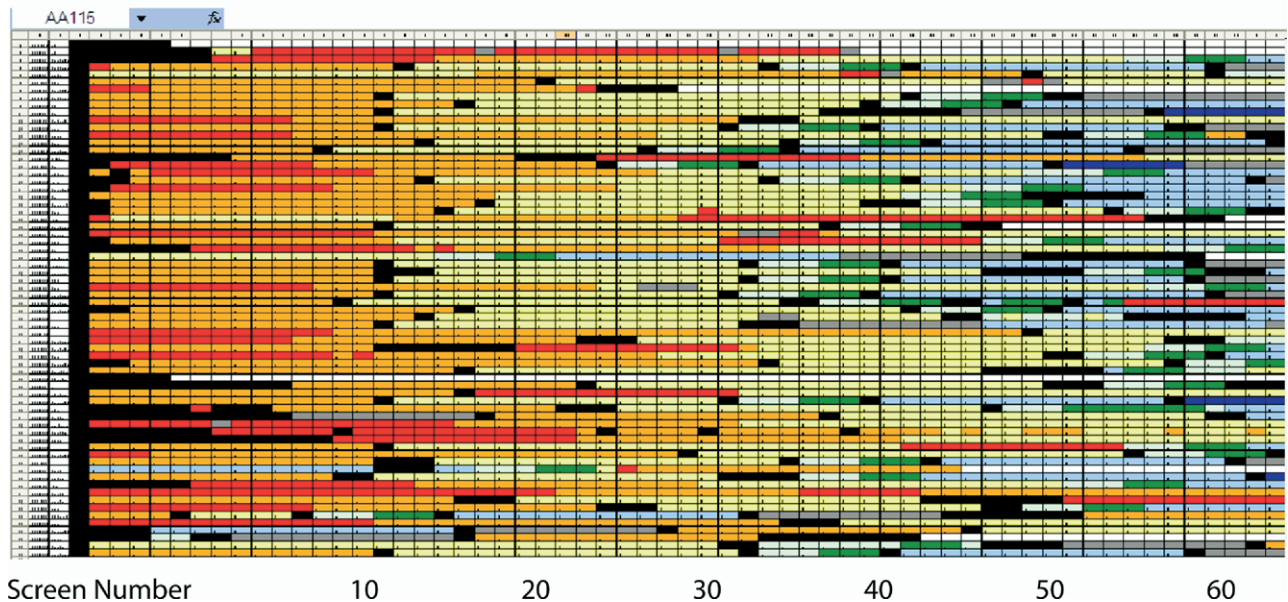


Figure 5. Visual representations of the path taken by the students through the tutorial. In the graphics above, each row represents the log file of one student, while each cell in the grid represents a single screen viewed in the computer tutorial. The cells are color-coded according to the legend given here, where a given color represents a specific chapter in the tutorial. Only the first 60 screens viewed are shown. (See Results for interpretation.)

Recent systematic reviews of medical computer-aided instruction confirm this impression (32–34).

A number of published editorials have suggested that more studies are needed where the comparison is not computer versus other medium but rather the instructional strategies that make computers most effective are uncovered (35, 36). This avoids a false “either/or” contrast and instead allows us to best blend these different modalities, using each for the particular context for which it is best suited (37).

We present a comparison of two different computer instructional strategies. Designers of radiology computer tutorials make decisions as to how they were going to layout their materials. At one end of the spectrum are strictly linear teacher-centered presentations where the content is arranged in a series of frames. This is akin to a lecturer working through a series of PowerPoint slides. The advantage of this layout is that it ensures that all of the material is presented using an organizing scheme that may help the student better encode the information (38). The disadvantage is that it does not allow the student to determine the direction of the instruction according to his or her own prior knowledge and learning needs (39).

At the other end of the spectrum is the random-access branched design, a good example of which is a reference Web site. Here the student can access any frame of material at any time, in any order. The student can create any path through the material (including a linear comprehensive one) according to his or her needs. The advantage is a customized learning experience. Also, active learning may occur during the creation of the path through the material (40). The disadvantage is that the extra cognitive work required to generate the path using the interface may make the process less efficient. In addition, the student may not generate a path that includes all of the material required for mastery of the domain (39).

We created two versions of a computer tutorial to teach medical students the skill of CSXR interpretation in the setting of trauma. Using a randomized block design, we found that the choice of layout did not affect the amount of learning but that the Web-layout group was considerably more efficient in learning.

Some students in the branched tutorial essentially converted the tutorial to a linear one. In these cases, the tutorial logs show a clear resemblance to the tutorial paths for the linear group. However, these cases were in the minority. Indeed, had it been the majority, we would have expected the branched group to take longer to complete the tutorial because they would have had to expend extra

cognitive effort to essentially “convert” the layout of the tutorial. In fact, the opposite was the case.

Students in the branched group were more efficient. They required considerably less time and fewer “screen views” to achieve a similar learning outcome. We noted qualitatively that the branched group spent less time on reviewing the basic anatomy of the cervical spine and on the section dealing with the alignment. They viewed the unknown images more often and spent more time with them.

Overall, both versions of the tutorial were effective in increasing the students’ skill, with the overall effect size being 2.1 with a 95% CI of 1.7–2.5. However, the students were still far from a mastery level, with the average student being able to correctly specify the nature of the lesion in only four of seven images. This is in line with published reports of medical student facility with radiological images. For example, Eisen et al. (41) found that medical students could interpret 8 of 20 chest radiographs in a test set correctly for which radiology residents correctly classified 18 of 20. On the other hand, in Lieberman et al.’s (27) study of a computer tutorial against a small-group tutorial, the students’ post-test scores in both groups were 7.7 correct of 8 images showing free air. It does not appear that normal images were included in their test, and no item analysis is included.

The students found it more difficult to accurately classify normal images than to spot abnormalities. In fact, the tutorial did not improve their ability to correctly classify normal images at all (effect size = -0.64). This may be because of the experimental situation, where, by completing a tutorial composed of only abnormal images, the students may be biased to find abnormalities. Bias has been shown to affect the ability of students to detect visual features (42). On the other hand, this may be due to the novice level of our subjects, as experts have been shown to distinguish themselves specifically in being able to rapidly discount normal findings (43).

There were several limitations to our study. The resolution of the images viewed by the students were certainly not at the level of a PACS system, being limited to a 1024×768 display on 17-inch computer monitors. This was true for both the intervention and the tests. Therefore, the improved skill of the students, and our numerical estimate of it, may not generalize to the real-world setting where more image information is available. Another threat to the generalizability of our findings was that we tested a relatively novice group of learners, as evidenced by their relatively low pretest scores. These

results may not apply to residents for example. While the vast majority of eligible students participated, there may have been an element of volunteer bias as we cannot exclude that students who did not participate are somehow different from our participants.

The reliability of our difference-score test was low at 0.21, although the overall reliability was good. While difference score reliabilities are known to be lower than corresponding reliabilities for single-test examinations, this increased measurement error increases the chances that we may have missed an effect of the layout that is indeed there (a Type II error). The wide CI around the effect size for the layout (-0.39 to $+0.27$) reflects this uncertainty. Ways to increase the reliability of our examination would be to increase the number of items and to decrease the heterogeneity of the items (i.e., exclude the normal examinations).

Our intervention comprised a specific mix of instructional strategies including the use of unknown images and visual segmentation with text explanations. While we believe that this mix of strategies is fairly representative of what a most instructional designers might use when faced with the same task, we cannot exclude the possibility that different choices of instructional strategies (say, increased use of repetitive practice) might have resulted in a different interaction with the tutorial layout variable that was our main manipulation.

CONCLUSIONS

The implications of our study for instructional designers concerned with the learning of radiology interpretation are the following:

- Both approaches appear to be equally efficacious in improving medical student ability to recognize radiographic features of abnormal CSXRs.
- Learner satisfaction is somewhat higher for the situation where the learning experience is highly programmed. We speculate it may be that students are reassured that they have not missed any significant points.
- In situations where the learner is time pressured (e.g., busy clinical rotations), it may be better to use a random access (Web) design. In this situation, the learners are considerably more efficient.
- Particular attention needs to be paid to demonstrating features of normal images that can be confused with

positive findings. Our tutorial was replete with images showing various abnormalities and an approach to systematically identifying those abnormalities. More examples of normal cases may have helped this deficiency.

ACKNOWLEDGMENTS

We thank the students at both centers for participating; Dr. Helen Nadel, the radiologist who independently re-read all of the images; and Dr. George Pachev for the pilot test item analysis.

References

1. Erkonen WE, D'Alessandro MP, Galvin JR, Albanese MA, Michaelsen VE. Longitudinal comparison of multimedia textbook instruction with a lecture in radiology education. *Acad Radiol* 1994; 1:287-292.
2. Jaffe CC, Lynch PJ. Computer-aided instruction in radiology: Opportunities for more effective learning. *AJR Am J Roentgenol* 1995; 164:463-467.
3. Skinner BF. *Science and Human Behavior*. New York: Macmillan, 1953.
4. Bloom BS. *Taxonomy of Educational Objectives: The Classification of Educational Goals*. Longmans, Green, New York; 1956.
5. Gagné RM, Briggs LJ. *Principles of instructional design*. New York: Holt, Rinehart and Winston, 1974.
6. Jonassen DH. Objectivism versus constructivism: Do we need a new philosophical paradigm? *Educ Technol Res Dev* 1991; 39:5-14.
7. Viccellio P, Simon H, Pressman BD, Shah MN, Mower WR, Hoffman JR. A prospective multicenter study of cervical spine injury in children. *Pediatrics* 2001; 108:e20.
8. Crocker L, Algina J. *Item Analysis. Introduction to Classical and Modern Test Theory*. Belmont, CA: Thomson Learning, 1986, pp 311-338.
9. Wright BD. Rack and stack: Time 1 versus time 2. *Rasch Measure Trans* 2003; 17:905-906.
10. Crocker LM, Algina J. *Procedures for Estimating Reliability. Introduction to Classical and Modern Test Theory*. Belmont, CA: Thomson Scientific, 1986, pp 131-156.
11. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Mahwah, NJ: Lawrence Erlbaum Associates; 1988.
12. Cumming G, Finch S. A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educ Psychol Measure* 2001; 61:532-574.
13. Kalb B, Gay SB. Internet resources for education in radiology. *Acad Radiol* 2003; 10(suppl 1):S81-S86.
14. Scarsbrook AF, Graham RN, Perriss RW. The scope of educational resources for radiologists on the internet. *Clin Radiol* 2005; 60:524-530.
15. ACR Campus. Available at <http://campus.acr.org/acr/Index.aspx>. Accessed April 24, 2007.
16. Aunt Minnie.com. Radiology Learning Center. Available at <http://www.auntminnie.com/index.asp?Sec=edu>. Accessed April 24, 2007.
17. Grunewald M, Heckemann RA, Wagner M, Bautz WA, Greess H. ELERA: A WWW application for evaluating and developing radiologic skills and knowledge. *Acad Radiol* 2004; 11:1381-1388.
18. Grunewald M, Heckemann RA, Gebhard H, Lell M, Bautz WA. COMPARE radiology: Creating an interactive Web-based training program for radiology with multimedia authoring software. *Acad Radiol* 2003; 10:543-553.
19. Angle JF, Gay SB, Hagspiel KD, Spinosa DJ, Leung DA, Matsumoto AH. Relative value of analog and digital teaching files. *Acad Radiol* 2002; 9:205-210.
20. Blunt D, O'Regan D. Using PACS as a teaching resource. *Br J Radiol* 2005; 78:483-484.
21. Scarsbrook AF, Foley PT, Perriss RW, Graham RN. Radiological digital teaching file development: An overview. *Clin Radiol* 2005; 60:831-837.

22. Weinberger E, Jakobovits R, Halsted M. MyPACS.net: A Web-based teaching file authoring tool. *AJR Am J Roentgenol* 2002; 179:579–582.
23. Bartlett ES, Maley JE, Fajardo LL. Radiology residency eCurriculum developed in-house: Evaluation of benefits and weaknesses. *Acad Radiol* 2003; 10:657–663.
24. Roubidoux MA, Chapman CM, Piontek ME. Development and evaluation of an interactive Web-based breast imaging game for medical students. *Acad Radiol* 2002; 9:1169–1178.
25. Chew FS, Relyea-Chew A. Distributed Web-supported radiology clerkship for the required clinical clerkship year of medical school: Development, implementation, and evaluation. *Acad Radiol* 2002; 9:713–720.
26. D'Alessandro MP, Galvin JR, Erkonen WE, Albanese MA, Michaelsen VE, Huntley JS, et al. The instructional effectiveness of a radiology multimedia textbook (HyperLung) versus a standard lecture. *Invest Radiol* 1993; 28:643–648.
27. Lieberman G, Abramson R, Volkan K, McArdle PJ. Tutor versus computer: A prospective comparison of interactive tutorial and computer-assisted instruction in radiology education. *Acad Radiol* 2002; 9:40–49.
28. Howerton WB Jr, Enrique PR, Ludlow JB, Tyndall DA. Interactive computer-assisted instruction vs. lecture format in dental education. *J Dent Hyg* 2004; 78:10.
29. Maleck M, Fischer MR, Kammer B, et al. Do computers teach better? A media comparison study for case-based teaching in radiology. *Radiographics* 2001; 21:1025–1032.
30. Collins J, Dotti SL, Albanese MA. Teaching radiology to medical students: an integrated approach. *Acad Radiol* 2002; 9:1046–1053.
31. Mileman PA, van den Hout WB, Sanderink GC. Randomized controlled trial of a computer-assisted learning program to improve caries detection from bitewing radiographs. *Dentomaxillofac Radiol* 2003; 32:116–123.
32. Greenhalgh T. Computer assisted learning in undergraduate medical education. *BMJ* 2001; 322:40–44.
33. Chumley-Jones HS, Dobbie A, Alford CL. Web-based learning: Sound educational method or hype? A review of the evaluation literature. *Acad Med* 2002; 77(10 suppl):S86–S93.
34. Letterie GS. Medical education as a science: The quality of evidence for computer-assisted instruction. *Am J Obstet Gynecol* 2003; 188:849–853.
35. Friedman CP. The research we should be doing. *Acad Med* 1994; 69: 455–457.
36. Cook DA. The research we still are not doing: An agenda for the study of computer-based learning. *Acad Med* 2005; 80:541–548.
37. Valcke M, De Wever B. Information and communication technologies in higher education: Evidence-based practices in medical education. *Med Teach* 2006; 28:40–48.
38. Mayer RE. Models for understanding. *Rev Educ Res* 1989; 59:43–64.
39. Large A. Hypertext instructional programs and learner control: A research review. *Educ Info* 1996; 14:95–106.
40. Schmidt RA, Bjork RA. New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychol Sci* 1992; 3:207–217.
41. Eisen LA, Berger JS, Hegde A, Schneider RF. Competency in chest radiography. A comparison of medical students, residents, and fellows. *J Gen Intern Med* 2006; 21:460–465.
42. Leblanc VR, Brooks LR, Norman GR. Believing is seeing: The influence of a diagnostic hypothesis on the interpretation of clinical features. *Acad Med* 2002; 77(10 suppl):S67–S69.
43. Myles-Worsley M, Johnston WA, Simons MA. The influence of expertise on x-ray image processing. *Learning Memory* 1988; 14:553–557.