**Title**
Expert-Novice Differences in Mammogram Interpretation

**Permalink**
https://escholarship.org/uc/item/9vs3q436

**Journal**
Proceedings of the Annual Meeting of the Cognitive Science Society, 29(29)

**ISSN**
1069-7977

**Authors**
Azevedo, Roger
Faremo, Sonia
Lajoie, Susanne P.

**Publication Date**
2007

Peer reviewed

# Expert-Novice Differences in Mammogram Interpretation

**Roger Azevedo**

Department of Psychology, Institute for Intelligent Systems, University of Memphis
3693 Norriswood Avenue, Memphis, TN, 38152, USA
{razevedo@memphis.edu}

**Sonia Faremo and Susanne P. Lajoie**
McGill University, Applied Cognitive Science Research Group,
3700 McTavish Street, Montréal, Quebéc, H3A 1Y2, Canada

## Abstract

This paper examines the results of two initial studies of the problem-solving strategies used by more and less skilled medical professionals during mammogram interpretation. The first study examined the cognitive processing of staff radiologists and radiology residents, while the second looked at surgical residents and medical students, as they individually solved a set of breast disease cases. Analyses of 100 verbal protocols from the two studies resulted in the development of a problem-solving model of mammogram interpretation and a characterization of novice expert differences based on performance measures. Results revealed that with increasing levels of expertise there were significant increases in the number of radiological observations and findings, proportion of correct diagnoses, use of data-driven problem solving, and diagnostic planning. The analysis provides a valuable initial characterization of mammogram interpretation across a broad range of expertise levels with implications for the design of computer-based learning environments aimed to train medical professionals to interpret mammograms.

**Keywords:** expertise; radiology; problem solving; diagnostic reasoning; think-aloud protocols; medical training

## Cognitive Science in the Real World: Improving Mammography Training Based on Expertise Studies

Breast cancer is the second leading cause of cancer deaths in women today after skin cancer (American Cancer Society [ACS], 2006). According to the World Health Organization (WHO), more than 1.2 million people will be diagnosed with breast cancer this year worldwide. Breast cancer is the leading cause of death among women 40 to 55 years of age and causes 18% of all cancer deaths in women. In 2007, an estimated 178,480 women in the US will be diagnosed with breast cancer and about 40,460 will die because of the disease (ACS, 2006). In the last decade, incidence rates stabilized probably because mammographic screening became a critical means of substantially reducing breast cancer mortality (ACS, 2000). Nevertheless, 11% to 25% of cancers are overlooked by radiologists on initial screening mammograms (Goergen et al., 1997).

Given the scope and seriousness of breast cancer, it is evident that any promising means for alleviating it should be investigated. Societal, ethical, and training issues should be investigated in order to lessen the impact of this disease. As cognitive scientists we have taken a critical initial step towards improving training in this area by conducting research that examines the cognitive components that constitute proficiency in mammogram interpretation. This paper links the results of two parallel studies in order to provide a comprehensive characterization of expert-novice differences in mammogram interpretation. Our intention is to subsequently use these research-based results to improve the training of future medical professionals (e.g., Azevedo & Lajoie, 1998; Lajoie & Azevedo, 2000; Crowley et al., 2005; Taylor, 2006).

### Cognitive Science Studies in Radiology

A few studies in radiology have been conducted by cognitive scientists focusing on the interpretation of chest x-rays. For example, Lesgold and colleagues (1981, 1988) provide one of the few existing explicit cognitive accounts of problem-solving strategies used by radiology residents and staff radiologists during chest x-ray interpretation. Their principal contribution was to demonstrate that experts extensively use "top-down" or "knowledge-based" processing. In this way, x-ray diagnosis is similar to but not the same as problem solving in non-perceptual domains. Other relevant research includes a study of chest radiography interpretation characterizing the interplay between perceptual and cognitive (knowledge-based) processing and a model of visual interaction (Rogers, 1992). This study identified three types of errors: (a) detection errors (failure to detect an abnormality), (b) labeling errors (mislabeling an abnormality), and (c) integration errors (correctly labeling an abnormality but failing to use it in the generation of a diagnostic hypothesis). Rogers was unable to examine expertise effects because of limited availability of participants in varying levels of expertise. A subsequent study by Raufaste and colleagues (1998) tested a model about how a human expert's cognitive system learns to detect, and does detect, pertinent data and hypotheses via a process called pertinence generation. Their results suggest two qualitative different kinds of expertise, basic and super. Basic experts are those who make routine daily diagnostic decisions by using all kinds of patient data (e.g., routine x-rays, clinical data) while super experts do the same but they also deal with

atypical cases due to their additional roles as clinical researchers who spend a tremendous amount of their career engaging in deliberate practice.

In sum, comparatively little cognitive research has investigated diagnostic radiology. The existing studies have provided initial characterizations of the diagnostic process, the role of schema-driven problem solving, and the top-down and bottom-up processes involved in diagnostic reasoning. They have also provided an initial understanding of the role of perceptual and hypothesis-driven processes. Again, these results have instructional implications that have not been widely used to inform the design of training.

## Objectives of the Study

In this study, the problem-solving strategies used by medical professionals with varying levels of training in mammogram interpretation were investigated. Three specific research objectives are addressed in this article. First, a model of problem solving in mammogram interpretation is presented, based on the analysis of verbal protocols. Second, the use of problem-solving strategies, operators, and control processes by participant groups is investigated. Third, the performance of participant groups differing on several measures (e.g. frequency and type of errors committed) is also investigated. In the discussion that follows, the results of this study are discussed in terms of how they can be applied to the design of training methods for medical professionals. The analyses are based on the amalgamation of two unpublished data sets.

## Method

The data from two unpublished parallel studies were analyzed in order to examine medical problem solving across a spectrum from novice to expert. Study 1 (Azevedo, 1998) examined staff radiologists and radiology residents, while study 2 (Faremo, 1997) looked at two less-experienced groups, medical students and surgical residents. Both studies were conducted within a teaching hospital system belonging to large private university in an eastern Canadian city, with the assistance of the same two medical experts (a surgeon and a radiologist both specializing in breast disease and mammography). The studies were parallel in terms of the research goals, experimental procedures, and data analysis techniques.

## Participants

A total of 36 participants drawn from the McGill University teaching hospitals took part in the two original studies. Study 1 included ten radiologists and ten radiology residents. Study 2 included eight undergraduate medical students and eight surgical residents. From these two groups five participants were randomly selected for the current study (a total of 10). The radiologists had MD degrees and Board Certification in radiology. The radiology residents and surgical residents had MD degrees and were on rotation at one of the teaching hospitals.

## Cases

Ten breast disease cases were used from the original studies. Similar cases were used in the two studies in accordance with the level of experience of the participants, disease categories, and mammographic manifestations. In both studies, an additional case was used as a practice case. Both medical professionals selected cases from their teaching files. Each case was comprised of a brief clinical history and at least four mammograms including the craniocaudal (CC) and mediolateral (MLO) views of the left and right breasts. For this study, a set of five cases was selected from each original study with the assistance of the consulting professionals. The cases included one benign and four malignant diseases (as confirmed by pathology reports). The cases also included common abnormalities as well as atypical ones that are infrequently encountered in mammography. Abnormalities ranged from ones that were fairly obvious to detect to those that required the use of a magnifying glass to detect.

## Experimental Procedure

The following description of the experimental procedure refers to both original studies. Participants were tested individually; the experimenter provided each participant with a one-page handout of instructions for the diagnostic task (e.g., *You will be presented with ten breast disease cases to diagnose. Each case will be comprised of a brief clinical history and a corresponding set of mammograms. For each case, please read the clinical history out loud, examine and describe the findings as you would normally. Suggest further examinations if appropriate. Please think out loud throughout the **entire** diagnostic process, that is, verbalize all comments and impressions you have as you diagnose each case.")*. S/he then placed the materials in front of the participant, including the practice case and the 10 cases. Each case was comprised of a manila envelope containing a typewritten clinical history and a set of mammograms. The experimenter presented each participant first with the practice case and subsequently with the 10 cases (order varied across participants). Video and audio data were collected during the entire experimental session. No time constraints were imposed.

## Analyzing the Think-Aloud Protocols

Audio and video data were transcribed according to the transcription conventions of Bracewell & Breuleux (1993) to ensure that the accuracy of lexical and syntactic structures was maintained as far as possible. The next section presents a detailed description of the coding scheme and the results of its application to all 100 transcribed and segmented protocols (five participants at four levels of expertise solving five cases). Segmented protocols and inter-rater reliability measures are provided.

**Coding Scheme**. Azevedo's (1997) coding scheme was based on three sources: (1) the content analysis, (2) theoretical and methodological articles (Chi, 1997; Ericsson, 2006; Ericsson & Simon, 1993), and (3) the results of previous studies in medical cognition (Hassebrock & Prietula,

1992; Patel & Ramoni, 1997), and chest radiography (Lesgold et al., 1981; 1988; Rogers, 1992).

The coding scheme consists of three major categories: *knowledge states*, *problem solving operators*, and *control processes* (Anderson & Labiere, 1998; Newell & Simon, 1972). Each of which is described below.

Knowledge States. Knowledge states include *radiological observations, radiological findings*, and *diagnoses* representing the hierarchical nature of medical knowledge in breast diseases and mammography (Evans & Gadd, 1989). *Radiological observations* are units of information that are recognized as potentially relevant in the problem-solving context (i.e., information from clinical histories and mammograms), but do not constitute clinically useful facts (e.g., presence of dense fibroglandular tissue on the mammograms). *Radiological findings* are units of information that are recognized as potentially relevant in the problem-solving context (i.e., information from clinical histories and mammograms) and which also constitute clinically useful facts (e.g., a cluster of pleomorphic calcifications on the mammograms). *Diagnoses* include disease types at different levels of abstraction, from pre-diagnostic labels to definitive diagnoses.

Problem-Solving Operators. Problem-solving operators are used to generate or instantiate states of radiological knowledge. Eleven basic types of operators were identified that characterize distinct segments of problem-solving behavior. They are inferred cognitive processes that modify, add, and/or eliminate existing or currently active knowledge states and produce new, active knowledge states. The operators reflect the knowledge and problem-solving behaviors required to successfully complete the diagnostic task. The conceptual operations involve actions that are or are not concurrently accompanied by verbalizations.

Control Processes. Control processes included *goals* (the use of the future tense to indicate an intended action), diagnostic *planning* (the planning of subsequent examinations and their possible interpretations), and *meta-reasoning* (a participant conducts a self-evaluation of the quality of the evolving diagnostic strategy).

## Inter-rater Reliability

Inter-rater reliability was established by recruiting a graduate student with experience in analyzing problem-solving transcripts. The student was trained to use Azevedo's (1997) coding scheme and was instructed to independently code the knowledge states, problem-solving operators and control processes from 60 randomly selected protocols (15 from each of the four groups). There was agreement on 446 out of a total of 472 coded segments (60 protocols with approximately eight segments each) yielding a reliability coefficient of .94 (Cohen's Kappa, $\kappa$ = .89). Inconsistencies were resolved through discussion between the experimenters and the student.

## Results and Discussion

Analysis of the 100 verbal protocols resulted in: (a) a problem-solving model of mammogram interpretation, and (b) a characterization of novice-expert differences related to this model. In this section we present the results of the inferential analyses that were conducted to verify whether there were any significant differences in the mean number of radiological findings, observations, and diagnoses across levels of expertise. In addition, non-parametric statistical analyses were conducted on the proportions (based on frequency data) of diagnostic accuracy, reasoning strategies, error types, requests for additional medical information, problem-solving operators, and control processes by level of expertise.

### Problem-Solving Model of Mammogram Interpretation

Solving a breast disease case involves examining and interpreting several sources of data in order to identify and characterize abnormalities and to arrive at diagnoses. The problem-solving model of mammogram interpretation decomposes this task into seven steps: (1) reading a clinical history, (2) placing a set of mammograms on a view-box and identifying individual mammograms in the set, (3) visually inspecting each of the mammograms, (4) identifying mammographic findings and observations, (5) characterizing mammographic findings and observations, (6) providing a definitive diagnosis or a set of differential diagnoses, and (7) specifying subsequent examinations (if required). These constitute a set of standard or general steps that are completed each time a practitioner diagnoses a breast disease case.

This model is consistent with how participants actually solved the cases, in that it allows for both a *linear approach* (e.g., from reading the clinical history to specifying subsequent examinations) and an *iterative approach* in which the results of a step may feed back to previous steps in the model. Using the linear approach (or data-driven problem solving) a participant reads the clinical history, scans the set of mammograms, identifies and characterizes the findings and/or observations, provides a diagnosis, and specifies a subsequent examination. The iterative approach (or mixed-problem solving strategy) involves some variation on the linear approach (e.g., a change in sequencing, repetition).

### Performance Measures
**Number of Radiological Findings, Observations, and Diagnoses.** Three one-way ANOVAs were performed on the mean number of *radiological findings*, *observations* and *diagnoses* across the four levels of expertise. The analyses revealed significant differences between the groups in the mean number of radiological observations ($F$ [3,16] = 9.98, $p$ < .05) and findings ($F$ [3,16] = 6.81, $p$ < .05). Post-hoc analyses failed to reveal significant differences based on the mean number of observations and findings between groups ($p$ > .05). There was no significant difference in the mean number of diagnoses between the groups ($F$ [3,16] = 2.54, $p$ > .05). On average, radiology residents and staff radiologists identified three observations per case while medical students

and surgical residents failed to identify any. For radiological findings, undergraduate medical students failed to identify any, but the other three groups identified at least one finding. All participants tended to provide approximately one diagnosis per case. The means and standard deviations for radiological observations, findings and diagnoses by level of expertise are presented in Table 1.

Table 1. Mean radiological observations, findings, and diagnoses by level of expertise.

| Performance Measures | Level of Expertise | | | |
|---|---|---|---|---|
| | Medical Students | Surgical Residents | Radiology Residents | Staff Radiologists |
| | Mean (SD) | Mean (SD) | Mean (SD) | Mean (SD) |
| Radiological Observations* | 0.16 (0.2) | 0.44 (0.3) | 2.56 (1.1) | 3.32 (1.9) |
| Radiological Findings* | 0.60 (0.2) | 1.36 (0.4) | 1.04 (0.3) | 1.16 (0.3) |
| Diagnoses | 0.92 (0.4) | 0.92 (0.2) | 1.36 (0.4) | 1.12 (0.1) |

Note: $^*p < .05$

Table 2. Proportion of diagnostic accuracy ratings, reasoning strategy, control processes, requests for additional medical information, and error types by level of expertise.

| | Level of Expertise | | | |
|---|---|---|---|---|
| | Medical Students | Surgical Residents | Radiology Residents | Staff Radio-logists |
| *Diagnostic Accuracy* * | | | | |
| Correct Diagnosis | .12 | .44 | .76 | .80 |
| Indeterminate Diagnosis | .16 | .36 | 0 | .08 |
| Wrong Diagnosis | .72 | .20 | .24 | .12 |
| *Reasoning Strategy* * | | | | |
| Hypothetico-Deductive | .68 | .32 | 0 | 0 |
| Data-Driven | .32 | .68 | .80 | .92 |
| Mixed | 0 | 0 | .20 | .08 |
| *Control Processes* * | | | | |
| Diagnostic Plans | .28 | .75 | .87 | .96 |
| Goals | .72 | .25 | .13 | .04 |
| *Error Types* * | | | | |
| Perceptual Detection | 0 | 0 | .83 | .60 |
| Wrong Recommendation | 0 | 0 | .17 | .40 |
| Multiple Errors | 1 | 1 | 0 | 0 |

Note: $^*p < .05$

**Diagnostic accuracy**. Diagnostic accuracy ratings take into account the combination of diagnoses and subsequent medical examinations. The two experts rated the final diagnosis provided in each case as *correct* (e.g., correct diagnosis and appropriate follow-up), *indeterminate* (e.g., a partially correct diagnosis with an inappropriate follow-up), or *wrong* (e.g., inappropriate follow-up for a diagnosis). A 3X4 Chi-square analysis revealed a significant difference in the distribution of the number of cases across levels of expertise and diagnostic accuracy ($\chi^2$ [6, $N = 100$] = 43.4, $p <$ .05) (see Table 2). Overall, staff radiologists and radiology residents provided significantly more correct diagnoses (80%

and 76%, respectively) than students or surgical residents (12% and 44%, respectively). In contrast, students and surgical residents provided significantly more incorrect diagnoses (72% and 20%, respectively) than staff radiologists and radiology residents (12% and 24%, respectively). Students and surgical residents also provided disproportionately more indeterminate diagnoses (16% and 36%, respectively) than the staff radiologists and radiology residents (8% and 0%, respectively).

The findings for most of the performance measures across the four levels of expertise are consistent with the expertise research in various domains. For example, across increasing levels of expertise there was a significant and consistent increase in the number of radiological observations and findings, and significant increases in the proportion of correct diagnoses, use of data-driven reasoning strategies and diagnostic planning. These results are consistent with certain robust findings in the expertise literature across domains (e.g., Feltovich et al., 2006; Norman et al., 2006)

The developmental trend in the results indicates that extensive medical training leads to organized knowledge structures, which in turn facilitate medical problem solving. The more experienced professionals were able to solve a higher proportion of cases using a data-driven reasoning strategy. They also engaged in extensive medical planning drawing on their organized knowledge bases to access meaningful patterns especially visual patterns. This led them to make an average of three observations, at least one finding and one diagnosis per case. In contrast, the less-experienced participants lacked the organized knowledge bases and corresponding access to meaningful patterns. As a result they could not elicit as many observations and findings and used mainly hypothetico-deductive reasoning, misdiagnosed a greater proportion of cases, and used more goal statements to support their hypothetico-deductive problem solving. Overall, participants provided on average one diagnosis per case. The two most experienced groups had learned to narrow their diagnoses to correct or suitable ones, while the two less experienced groups were not able to do so and may not even have known many of the disease types encountered in mammography. These findings may also be explained by the fact that mammography is a well-constrained sub-specialty of radiology. Further, the levels of abstraction in diagnostic hypotheses are not considered important in mammography, which may also have contributed to similar performance (i.e., average number of diagnosis) between the groups.

**Problem-Solving Strategies**. Each protocol was categorized in terms of predominant problem solving strategy. The types were: (1) hypothetico-deductive, a form of backward problem solving involving hypothesis generation, information search, data interpretation and hypothesis evaluation; (2) data-driven, where one proceeds from reading the clinical history to specifying subsequent examinations; and (3) mixed-strategy, a combination of data-driven and goal-driven problem solving strategies. A 3X4 Chi-square analysis revealed a significant difference in distribution of strategies used across

levels of expertise ($\chi^2$ [6, $N = 100$] = 48.1, p < .05; see Table 2). Overall, the medical students diagnosed the cases using mainly hypothetico-deductive reasoning (68%) but sometimes used a data-driven strategy (32%). In contrast, the surgical residents used mainly data-driven (68%) and rarely used hypothetico-deductive reasoning (32%). As for the two more-experienced groups, they both tended to use the data-driven strategy (80% and 92%, respectively) and sometimes used a mixed-strategy (8% and 25%, respectively).

The proportion of problem-solving strategy types used also differed based on the level of expertise. The two less experienced groups used hypothetico-deductive reasoning while the two more experienced groups did not use the strategy at all. In contrast, the more experienced groups used a mixed reasoning strategy while the two least experienced groups did not use it all. There was an increase in the use of the data-driven strategy with increasing expertise. As previously discussed, the results are consistent with previous research that has shown that the extensive knowledge of experts permits rapid recognition and rapid schema triggering possibly at the expense of problem understanding and problem-solving search (e.g., Lesgold et al., 1981, 1988). This provides an explanation for the increasing use of data-driven reasoning strategies with increasing levels of expertise. It also provides an explanation for why the two least experienced groups used hypothetico-deductive reasoning — they lacked a coherent, interconnected knowledge base that would permit them to use data-driven reasoning. Instead, they reasoned backwards by engaging in hypothesis generation, information search, data interpretation and hypothesis evaluation.

The use of a mixed strategy solely by the two more experienced groups is particularly interesting and has several cognitive and training implications. First, it suggests they used their extensive, highly-organized knowledge bases in a data-driven mode until it was no longer advantageous and then reverted to a goal-driven strategy. The reversal from data-driven to goal-driven relates to findings from the expertise literature which shows that experts have superior self-monitoring skills and self-knowledge skills. As noted, expertise research dealing specifically with the development or use of metacognitive skills is lacking. We propose that after experts attempt to use their knowledge base to interpret and solve a case, they then frame goals, select tactics and/or strategies which they predict can be used successfully to reach those goals, they then apply the tactics or strategies and observe the results. This ability to self-regulate may be based on their understanding of the limits of their knowledge base. However, they are strategic in setting goals which they are likely to reach (i.e., providing an accurate solution).

**Frequency of Control Process Use**. Regardless of level of expertise participants used two main control processes, diagnostic plans and goals. A 2X4 Chi-square analysis revealed a significant difference in the distribution of control processes used across levels of expertise ($\chi^2$ [3, $N = 138$] =

29.1, p < .05; see Table 2). Overall, surgical residents, radiology residents, and radiologists tended to use more diagnostic plans (75%, 87%, and 96% of the cases, respectively) than medical students (28% of the cases). In contrast, medical students tended to use more goals (72% of the cases) than surgical residents, radiology residents, and radiologists (25%, 13%, and 4% of the cases, respectively).

**Types of Errors Committed During Diagnostic Reasoning.** An analysis of the 46 errors (on 100 cases) committed by the participants revealed three major types: (1) perceptual detection errors (failure to detect a finding), (2) wrong recommendation errors (proposing an inappropriate subsequent examination), and (3) multiple errors (combination of perceptual detection, finding mischaracterization, no diagnosis, wrong diagnosis or wrong recommendation). A 3X4 Chi-square analysis revealed a significant difference in distribution of error types across levels of expertise ($\chi^2$ [3, $N = 46$] = 34, p < .05; see Table 2). Overall, medical students and surgical residents committed more errors (88% and 52% error rates, respectively) than the radiology residents or staff radiologists (20% and 12% error rates, respectively). A further analysis of the errors revealed that the two less-experienced groups committed multiple errors while the two more-experienced groups committed single errors only (either perceptual detection or wrong recommendation).

The errors committed by the participants can be analyzed based on level of expertise and the number of errors committed while solving a case. The more experienced professionals typically committed one error, either a perceptual detection error or a wrong recommendation error. The few perceptual detection errors committed can be explained by one of the pitfalls of being an expert — the rapid instantiation of a schema based on an extensive organized knowledge base leads to an incomplete extraction of meaningful patterns in the data. As noted earlier, this problem is widely documented in the expertise literature and leads to a trade-off between speed and accuracy (Feltovich, Spiro, & Coulson, 1997).

## Conclusions

In conclusion, we believe this study provides a valuable initial characterization of mammogram interpretation across a broad range of expertise levels. In addition, it contributes to the wealth of existing expertise studies in non-visual medical domains (e.g., Norman et al., 2006). The results have provided a research base from which we have derived training implications for medical professionals (Crowley et al., 2005; Lajoie & Azevedo, 2000; Taylor, 2006). We propose that future work in this area should focus on building a more comprehensive model of the perceptual and cognitive processes underlying mammogram interpretation and determining the implications for training. This may best be accomplished by drawing on various theoretical perspectives and incorporating the results of various types of research. For example, researchers with converging theoretical and

methodological orientations may contribute to our understanding of radiological expertise by conducting (1) studies of reaction times to assess detection abilities, (2) fMRI studies to examine the role of cortical structures during mammogram interpretation, (3) longitudinal studies to assess the quantitative and qualitative changes of emerging knowledge structures and problem solving strategies during the course of one's medical training, and (4) conversational and gestural analyses of teaching rounds focusing on how staff radiologists frame tutoring sessions, ask questions, aid students during problem solving, and react to student errors (verbally and non-verbally). In sum, future research endeavors should continue the effort to further our understanding of the interaction between perceptual and cognitive factors underlying mammogram interpretation and to improve future radiological training.

## References

Anderson, J. & Labiere, C. (1998). *Atomic components of thought*. Hillsdale, NJ: Erlbaum.

Azevedo, R. (1997). *Expert problem solving in mammogram interpretation: A visual cognitive task*. Unpublished doctoral dissertation, McGill University, Montréal, Canada.

Azevedo, R., & Lajoie, S. (1998). The cognitive basis for the design of a mammography interpretation tutor. *International Journal of Artificial Intelligence in Education, 9*, 32-44.

Bracewell, R., & Breuleux, A. (1993). Substance and romance in analyzing think-aloud protocols. In P. Smagorinsky (Ed.), *Speaking about writing: Reflections on research methodology* (pp. 55-88). Newbury Park, CA: Sage.

Chi, M. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences, 6*, 271-315.

Cooper, R. (1992). Mammography. In J. Isaacs (Ed.), *Textbook of breast disease* (pp. 47-86). St. Louis, MO: Mosby.

Crowley, R., Naus, G., Stewart, J., & Friedman, C. (2003). Development of visual diagnostic expertise in pathology - An information-processing study. *Journal of the American Medical Informatics Association, 10*, 39-51.

Ericsson, K. (2006). Protocol analysis and expert thought: Concurrent verbalizations of thinking during experts' performance on representative tasks. In K. Ericsson, N. Charness, P. Feltovich, & R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 223-241). Cambridge, MA: Cambridge Press.

Ericsson, K., & Simon, H. (1993). *Protocol analysis: Verbal reports as data)*. Cambridge, MA: MIT Press.

Evans, D., & Gadd, C. (1989). Managing coherence and context in medical problem-solving discourse. In D. Evans & V.Patel (Eds.), *Cognitive science in medicine: Biomedical modeling* (pp. 211-255). Cambridge, MA: MIT Press.

Faremo, S. (1997). *Novice diagnostic reasoning in a visual medical domain: Implications for the design of a computer-based instructional system for undergraduate medical education*. Unpublished master's thesis, Concordia University, Montréal, Canada.

Goergen, S., Evans, J., Cohen, G., & MacMillan, J. (1997). Characteristics of breast carcinomas missed by screening radiologists. *Radiology, 204*, 131-135.

Hassebrock, F., & Prietula, M. (1992). A protocol-based coding scheme for the analysis of medical reasoning. *International Journal of Man-Machine Studies, 37*, 613-52.

Lajoie, S., & Azevedo, R. (2000). Cognitive tools for medical informatics. In S.P. Lajoie (Ed.), *Computers as cognitive tools II: No more walls: Theory change, paradigm shifts and their influence on the use of computers for instructional purposes* (pp. 247-271). Mahwah, NJ: Erlbaum.

Lesgold, A., Feltovich, P., Glaser, R., & Wang, Y. (1981). *The acquisition of perceptual diagnostic skill in radiology* (Tech. Rep. No. PDS-1). Pittsburgh: University of Pittsburgh, Learning Research and Development Center.

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing x-ray pictures. In M. Chi, R. Glaser & M. Farr (Eds.), *The nature of expertise* (pp. 311-342). Hillsdale, NJ: Erlbaum.

Newell, A., & Simon, H. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.

Norman, G., Eva, K., Brooks, L., & Hamstra, S. (2006). Expertise in medicine and surgery. In K. Ericsson, N. Charness, P. Feltovich, & R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performanc*e (pp. 339-353). Cambridge, MA: Cambridge Press.

Patel, V., & Ramoni, M. (1997). Cognitive models of directional inference in expert medical reasoning. In P. Feltovich, K. Ford & R. Hoffman (Eds.), *Expertise in context: Human and machine* (pp.67-99). Menlo Park, CA: AAAI Press.

Raufaste, E., Eyrolle, H., & Marine, C. (1998). Pertinence generation in radiological diagnosis: Spreading activation and the nature of expertise. *Cognitive Science, 22*, 517-546.

Rogers, E. (1992). *Visual interaction: A link between perception and problem solving*. Unpublished dissertation, Georgia Institute of Technology, Atlanta, GA.

Taylor, P. (2006). *From patient data to medical knowledge: Principles and practice of health informatics*. London, UK: Blackwell.