The Effects of Disease Category on Diagnostic Problem Solving in Mammography

Roger Azevedo [1], Gwyneth Lewis [1], Roberta Klatzky [2] and Emily Siler [1]

{razevedo, glewis1, easiler@memphis.edu}
Department of Psychology / Institute for Intelligent Systems
400 Innovation Drive, Memphis, TN, 38152, USA

{klatzky@cmu.edu}
Department of Psychology, Carnegie Mellon University
5000 Forbes Avenue, Baker Hall, Pittsburgh, PA, 15213, USA

Abstract

Research on diagnostic problem solving is devoted to understanding the cognitive factors underlying the superior performance exhibited by medical professionals in their domain. In this study we analyzed think-aloud protocols and video recordings of staff and resident radiologists as they diagnosed mammograms during an interactive problem solving task. Analyses revealed statistically significant differences in radiologists' usage of problem solving operators (PSOs) and several performance measures between benign and malignant breast disease cases. Results suggest that additional factors outside of expertise should be considered in order to better understand diagnostic problem solving in a medical domain. The effects of disease type on performance and diagnostic problem solving should guide future development of medical tutoring systems.

Keywords: problem solving, expertise, medicine, think-aloud protocols, diagnosis, intelligent tutoring systems

Introduction

Research on medical expertise has broadened our understanding of the cognitive processes involved in medical problem solving. Much of our understanding of medical expertise is based on findings from medical text comprehension studies and existing expertise studies in non-visual medical domains (e.g., Norman et al., 2006). Relatively little is known about the functioning of expertise during visual medical tasks (Azevedo, 1997, Azevedo et al., 2007; Faremo, 1997; Lesgold et al., 1981, 1988; Rogers, 1992), such as interpreting and diagnosing x-rays in mammography.

Little emphasis is placed on the role of case type on performance and the diagnostic process. Previous research has documented the effect of case difficulty on clinical reasoning (Brooks, Norman, & Allen, 1991). Although experts are distinguished from novices by their forward reasoning approach, this is only true for more difficult cases. Experts tend to rely on pattern recognition when solving easier cases. Other factors impacting diagnostic problem solving have included the order of data presentation (Bergus, Chapman, Gjerde, & Elstein, 1995), the amount of detail presented in a case description (Redelmeir, Koehler, Liberman, & Tversky, 1995), and familiarity of the material (Bordage, 1999). The present study contributes to an important area of research by examining the impact of disease category on radiologists' mammogram interpretation during a diagnostic problem solving task. Disease category refers to ultimate diagnostic outcome, (i.e., benign or malignant).

Breast Disease

Mammography is considered the best means of detecting breast cancer in its early stages because cancer can develop long before any symptoms are felt. Abnormal mammograms constitute about 5%-10% of routine screenings. Of this percentage, most are diagnosed as benign, meaning the breast tissue is not cancerous. Unfortunately, the error rate in mammography is high. An estimated 10%-20% of breast cancers are missed in routine screenings (American Cancer Society [ACS], 2007). According to the Physician Insurers Association of America (2002), radiology is the target of more malpractice suits (33%) than any other specialty with the top three reasons involving misreading mammograms, underestimating seriousness of findings, and failing to detect findings.

Serious training issues in the area of mammography require attention. The importance of understanding problem solving in this context is profound because radiologists are faced with detecting potentially life threatening diseases on a daily basis. 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 presents the results of a study designed to build a comprehensive characterization of differences in mammogram interpretation and the impact of disease category on the diagnostic problem solving processes of mammographers. However, space limitations preclude us from providing detailed analyses about how to design medical training systems, based on the results of this study.

Method

Participants

A total of 13 participants drawn from the University of Pittsburgh Medical Center teaching hospitals took part in the study including nine staff radiologists and four radiology residents. The radiologists had MD degrees and Board Certification in radiology. The radiology residents also had

MD degrees but were on rotation at one of the teaching hospitals. In this study, we decided to combine the two groups of medical professional into one group given the following constraints: (1) small sample size in each group, (2) limited ability to conduct statistical analyses given (1), and (3) our main interest in examining the role of disease category on diagnostic problem solving and other performance measures.

Benign and Malignant Breast Disease Cases

Ten breast disease cases were used with an additional case used for practice. A staff radiologist worked with the team and selected cases from her teaching files, none of which were previously viewed by the participants. She served as the consulting expert and did not participate in the experiment. Each case comprised a brief clinical history and four mammograms including the craniocaudal (CC) and mediolateral (MLO) views of the left and right breasts and any additional data necessary to provide a complete diagnosis (e.g., spot compression, ultrasound scan, previous mammograms, etc.). For this study, a set of 10 cases was selected that included five benign and five 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 (e.g., large mass) to those that required the use of a magnifying glass to detect (e.g., cluster of microcalcifications).

Experimental Procedure

Participants were tested individually; the experimenter provided each participant with a one-page handout of instructions for the think-aloud diagnostic task (i.e., *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. Please let me know if you would like to review additional data during the diagnostic process*"). The experimenter then presented each participant with the same cases selected from the consulting expert's teaching files (i.e., first with the practice case and subsequently with the 10 cases with order varied across participants). Participants could request additional data from the experimenter (e.g. an ultrasound image) during the diagnostic task. 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 by a trained research assistant according to the transcription conventions of Bracewell & Breuleux (1993) to ensure accuracy of lexical and syntactic structures. Overall protocol length for all participants was about 76 pages of single spaced text (about 6 pages per participant) and total word length was over 33,000 words (average of 2,538.5 words per participant). The next section presents a detailed description of the coding scheme and the results of its application to all 130 transcribed and segmented protocols (13 participants solving 10 cases). Segmented protocols and inter-rater reliability measures are provided.

Coding Scheme Azevedo and colleagues' (1997, 2007) coding scheme was used for the current analysis with a few additions. The coding scheme is based on three sources: (1) the content analysis of mammography, (2) theoretical and methodological articles (Chi, 1997, 2006a, 2006b; Ericsson, 2006; Ericsson & Simon, 1993), and (3) the results of previous studies in medical cognition (Hassebrock & Prietula, 1992; Patel & Ramoni, 1997), and chest radiography (Faremo, 1997; Lesgold et al., 1981, 1988; Rogers, 1992). Azevedo's (1997) coding scheme consists of three major categories: *knowledge states*, *problem solving operators*, and *control processes* (Anderson & Labiere, 1998; Newell & Simon, 1972). All are 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 (e.g., malignant with in situ ductal carcinoma).

Problem-solving operators (PSOs). 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. For this study, a new operator termed *Data Requisition* was created to supplement the original coding scheme as this study included an additional interactive component involving soliciting the experimenter for more data.

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). For the current study, *Goal* was categorized as a type of problem-solving operator for the current analysis (see previous paragraph).

Application of the Coding Scheme The consulting radiologist was video recorded while solving each of 11 cases (one practice case and 10 additional cases). A research assistant with training in Azevedo's (1997) coding scheme independently coded the 130 protocols (13 participants x 10 cases) for problem solving operators. The experimenter re-coded 70% of the protocols to establish inter-rater reliability. Agreement surpassed 97% (1,117 out 1,146 PSOs) and disagreement was resolved through regular discussion. The observations, findings, and diagnoses gleaned from the video data served as the gold standard by which participant observations were compared for accuracy. Problem-solving operators were coded using an augmented version of Azevedo and colleagues' (1997, 2007) coding scheme. All of the measures, including incidental reading and scan time, are described below.

Diagnoses. Diagnoses were coded as either *correct* or *incorrect* based on their consistency with the consulting expert's diagnoses. *Correct* diagnoses matched the consulting expert's diagnosis, even if at different levels of abstraction (e.g., *carcinoma* and *suspicious for malignancy*) and was given a score of 1 for a match with the consultant's diagnosis. For example, a participant's diagnosis of infiltrating ductal carcinoma matches the consultant's diagnosis for that particular case. *Incorrect* diagnoses did not match the consultant's diagnosis and was given a score of 0. For example, a participant gave a diagnosis of microcalcifications while the correct diagnosis provided by the consultant for the same case was benign mass. It should be noted that the exact diagnoses provided by the consultant are based on postmortem pathology reports.

Observations. Observations were identified in the protocols for each participant. An observation was given a score of 1 if it closely matched the gold standard (e.g., the observation *heterogeneously dense tissue* closely matches *rather dense tissue in both breasts*) and 0 if it did not.

Findings. Similar to observations, findings were identified in the protocols and compared with the findings of the consulting radiologist. Correct findings were given a score of 1 if they closely matched the gold standard set by the consulting radiologist and 0 if they did not.

No codes. Protocols containing utterances that did not fit with any of the operational problem-solving definitions were coded as *no code* or rare instances when participant speech was inaudible. *No codes* comprised 11% of the data and were excluded from the analysis.

Reading time. Reading time was the total time in seconds a participant spent reading the clinical case history. A component of scanning time, these data were obtained from the time stamp on the video and the session.

Scanning time. Scanning time was coded as the time in seconds a participant took to remove the mammograms from the envelope, arrange the mammograms on the viewbox, and inspect the mammograms with the naked eye and/or magnifying glass up until the participant made their first utterance. This data was obtained from the time stamp on the video and the session.

Results

The data were analyzed to determine the effects of disease category (benign, malignant) on participants' diagnostic performance. Inferential statistics were conducted on the proportions (based on frequency data) of problem-solving operators and performance measures.

Performance Measures

Several ANOVAs were performed on participants' performance data, including scanning time, reading time, findings and observations accuracy, and diagnostic accuracy to determine the effects of disease category. The performance data of each participant were expressed as proportions under case type. Results revealed statistically significant results, based on disease category, for findings and observations accuracy and diagnostic accuracy (see Table 1). Participants accurately detected significantly more observations and findings on malignant cases. They also provided significantly more correct diagnoses for benign diseases (i.e., misses exceeded false alarms). Overall, participants did not significantly differ in the amount of time to construct an initial representation of a case or the overall amount of time necessary to diagnose each case.

Table 1: Accuracy measures by disease category.

Performance Measure	Disease Category	
	Benign (n=5)	Malignant (n=5)
	Mean (SD)	Mean (SD)
Mean scanning time (sec.)	46.12(23.14)	51.65(34.25)
Mean reading time (sec.)	206.28(99.2)	234.53(105.80)
Findings & observations accuracy*	0.61(0.46)	0.73(0.29)
Diagnostic accuracy*	0.62(0.21)	0.53(0.30)

Note: * $p < .05$

Problem Solving Operators (PSOs)

The coded problem solving operators were tallied for all cases under the categories of PSO type, participant number, and case number, and subsequently converted into proportions by case and participant. Individual ANOVAs were performed on

each PSO, by disease category (benign, malignant). Results revealed significant differences ($p<.05$) for five PSOs, including *Data Classification, Data Explanation, Data Requisition, Hypothesis Evaluation,* and *Hypothesis Generation*. Details are discussed below in terms of malignancy type (see Table 2 for a full list of the proportional problem solving operator usage in benign and malignant cases). Additionally, the results yielded a model of the diagnostic process involving the order of PSO application.

Table 2: Mean proportional problem solving operator usage based on disease category.

Operator	Disease Category	
	Benign (n=5) Mean(SD)	Malignant (n=5) Mean(SD)
Data Acquisition	0.10 (0.05)	0.10 (0.06)
Data Assessment	0.04 (0.07)	0.02 (0.04)
Data Classification*	0.07 (0.07)	0.04 (0.06)
Data Comparison	0.03 (0.04)	0.04 (0.05)
Data Examination	0.19 (0.10)	0.19 (0.10)
Data Explanation*	0.06 (0.07)	0.02 (0.04)
Data Exploration	0.18 (0.11)	0.14 (0.09)
Data Identification	0.01 (0.04)	0.02 (0.04)
Data Requisition*	0.17 (0.10)	0.13 (0.07)
Goal	0.03 (0.06)	0.02 (0.04)
Hypothesis Evaluation*	0.01 (0.04)	0.03 (0.05)
Hypothesis Generation*	0.01 (0.02)	0.12 (0.09)
Recommendation	0.07 (0.07)	0.07 (0.05)
Summarization	0.04 (0.06)	0.04 (0.07)

Note: * $p < .05$

Benign Cases. Significantly greater proportions of the operators *Data Classification, Data Explanation,* and *Data Requisition* were observed in benign cases. The operator *Data Classification* typically occurs after mammographic features (findings and observations) have been identified, characterized, and/or compared with other mammographic views. This result may be interpreted as due to the importance of detecting all potentially relevant features. Mammograms with benign diagnoses may thus entail a more exhaustive search, possibly involving iterative assessments in order to satisfactorily rule out disease. *Data Explanation* is associated with interpreting the significance of mammographic data (e.g., clinical history, findings) in terms of the likely underlying physiological cause. Greater usage of this operator when solving benign cases may, as previously hypothesized, be due to the ambiguity in solving benign cases. Diagnosing benign mammograms may involve greater iteration between states than when diagnosing malignant mammograms, which may be due to the crucial nature of identifying potentially significant or clinically relevant mammographic features. The assumption is that diagnostic uncertainty in benign cases leads to a wider range of potential pathophysiological causes. *Data Requisition* involves soliciting the experimenter for additional data (e.g., spot compression views, ultrasound image) to compare or confirm the existence, location, or appearance of mammographic features between or across

views. Additionally, this operator may be used to test hypotheses about the potential cause or diagnosis either to confirm the location or existence of a suspected finding. Higher prevalence of this operator for benign cases may again be due to the iterative nature in solving benign cases.

Malignant Cases. Significantly greater proportions of *Hypothesis Evaluation* and *Hypothesis Generation* were observed in malignant cases. *Hypothesis Evaluation* occurs after a hypothesis has been generated. This is a sort of "follow-up" procedure used to assess how well mammographic features measure support a generated hypothesis. This may entail confirmation or rejection of a hypothesis (which may lead to the generation of alternative hypotheses). There may be a higher proportion of this operator for malignant cases simply because there is a greater proportion of hypothesis generation. *Hypothesis Generation* is typically observed in the later stages of mammography diagnosis as it frequently occurs only after mammographic features have been identified, characterized, and classified. Typically, a hypothesis consists of a preliminary statement regarding a mammogram's diagnosis. The operator may be more common in solving malignant cases because malignant cases require less iteration between the earlier problem solving states. Diagnostic training is assumed to focus on malignant pattern recognition, whereas the recognition of benign features may be less emphasized.

Problem Solving Model Based on Disease Type
The study yielded an interpretive model of the problem solving stages and a representation of the associated problem solving operators by disease type. The number of instances associated with this model will not be discussed for the purposes of this paper. The process for benign-case solving (BCS) and the process for malignant-case solving (MCS) initially follow the same linear sequence during the early stages of the problem solving process. These stages are as follows:

First, the clinical patient history is read to acquire patient information (*Data Acquisition*) that provides the basis for the next step of requesting additional data or mammograms (*Data Requisition*) followed by positioning the data on the viewbox and assessing the technical quality (*Data Assessment)*. A goal may then be verbalized (*Goal)* that is pursued through a visual scan of the data to identify relevant mammographic features (*Data Examination)*. The features are then characterized on the basis of their visual properties (*Data Exploration)* that determine the next step of classifying the data (*Data Classification)*.

At this point, the BCS process follows a different path than the MCS process. The BCS process follows *Data Classification* by stating the pathophysiological cause (*Data Explanation*) and then iterates back to *Data Requisition* followed by the linear steps of placing, identifying, scanning, examining, exploring, and classifying the new data. During these subsequent steps, goals may be verbalized *(Goal)* and pursued by comparing the new data with previous data (*Data*

Comparison) that will lead to further classification and explanation of the data.
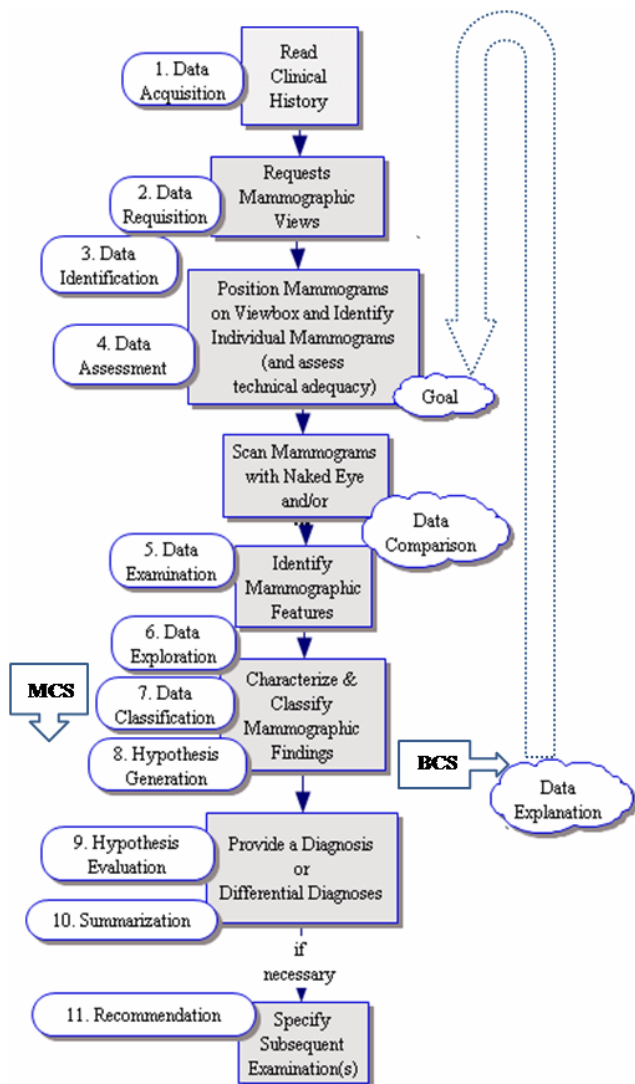

Figure 1. Problem solving model for benign (BCS) and malignant (MCS) breast disease cases.

Additionally, the process may again iterate to *Data Requisition* and repeat. The MCS process, however, proceeds to generating a hypothesis to explain the classified feature (*Hypothesis Generation)*. The hypothesis is then tested and evaluated (*Hypothesis Evaluation*), which if rejected may lead to further instances of *Hypothesis Generation.* The BCS process eventually progresses to *Hypothesis Generation* and *Hypothesis Evaluation* after some iteration. Final steps in the process involve summarizing the relevant information *(Summarization)* and specifying treatment or follow-up procedures *(Data Recommendation)* if necessary.

Conclusion

The results contribute to an emerging body of research aimed at understanding the cognitive factors underlying diagnostic problem solving of breast disease cases and the differences in diagnostic accuracy based on disease category. Future research may aim towards better understanding of the role of medical data in diagnostic problem solving. Our study suggests that findings and observations are far less often detected in malignant cases than in benign cases. This could be attributed to mammographers' awareness of the high stakes involved in diagnosing mammograms. Mammographers may spend more time on benign cases to ensure that all potentially malignant features are detected. Understanding of detection differences when solving benign and malignant cases may benefit from eye tracking studies to determine whether case type is responsible for detection differences or simply reporting differences. Our study also suggests that diagnostic accuracy is higher for malignant cases. Are malignant cases simply easier to solve? This is a difficult question considering that detection of mammographic features was higher for benign cases. These questions demonstrate a need for future research in this area.

Future research should strive to have researchers from various fields of cognitive science with divergent theoretical and methodological backgrounds contribute to our understanding of disease type in medical problem solving by conducting: (1) eye tracking studies to assess detection of benign and malignant features and therefore examine search processes and visual chunking and their relation to problem solving processes and diagnostic accuracy, (2) longitudinal studies to assess the development of expertise in medical problem solving skills and expertise, (3) conversational and gestural analyses of teaching rounds focusing on how staff radiologists train medical students and residents to problem solve, and how they provide different types of scaffolding to correct different types of performance issues (e.g., strategies for reducing differential diagnoses), and (4) utilize these performance data and the problem solving model to re-design medical training systems. In sum, future research should continue the effort to further our understanding of mammogram interpretation of benign and malignant cases.

Acknowledgments

References

American Cancer Society (ACS). *Breast Cancer Facts & Figures 2007-2008*. Atlanta: American Cancer Society, Inc.

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

Azevedo, R., Faremo, S., & Lajoie, S.P. (August, 2007). *Expert-novice differences in mammogram interpretation*. In D. S.

McNamara & J. G. Trafton (Eds.), Proceedings of the 29th Annual Cognitive Science Society (pp. 65-70). Austin, TX: Cognitive Science Society.

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

Azevedo, R. (1998). Expert problem solving in a visual medical domain. In M. A. Gernsbacher, & S. J. Derry (Eds.), Proceedings of the Twentieth Annual Conference of the Cognitive Science Society (pp. 72-77). Mahwah,NJ: Erlbaum.

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.

Azevedo, R., Faremo, S., & Lajoie, S. P. (August, 2007). *Expert-novice differences in mammogram interpretation*. In D. S. McNamara & J. G. Trafton (Eds.), Proceedings of the 29th Annual Cognitive Science Society (pp. 65-70). Austin, TX: Cognitive Science Society.

Bergus, G. R., Chapman, G. B., Gjerde, C., & Elstein, A. S. (1995). Clinical reasoning about new symptoms despite preexisting disease: Sources of error and order effects. *Family Medicine, 27*, 314-320.

Bordage, G. (1999). Why did I miss the diagnosis? Some cognitive explanations and educational implications. *Academic Medicine, 74*, 8138-8143.

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.

Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). The Role of Specific Similarity in a Medical Diagnostic Task. *Journal of Experimental Psychology: General, 120,* 278-287.

Carney, P., Yi, J., Abraham, L., & Miglioretti, D. (2007). Reactions to uncertainty and the accuracy of diagnostic mammography. *Society of General Internal Medicine*, *22*, 234-241.

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

Chi, M. T. H. (2006a). Two approaches to the study of experts' characteristics. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 21–30). Cambridge: Cambridge University Press.

Chi, M. T. H. (2006b). Laboratory methods for assessing experts' and novices' knowledge. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 167–184). Cambridge: Cambridge University Press.

Crowley R S, Tseytlin E, Jukic D. ReportTutor - An intelligent tutoring system that uses a natural language interface. *AMIA Annual Symposium Proceedings* (2005), 171-5.

Crowley, R. S., Legowski, E., Medvedeva, O., Tseytlin, E., Roh, E., & Jukic, D. (2007). Evaluation of an intelligent tutoring system in pathology: Effects of external representation on performance gains, metacognition, and acceptance. *Journal of the American Medical Informatics Association, 14,* 182-190.

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.

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.

Redelmeier, D. A., Koehler, D.J., Liberman, T. V., & Tversky A. (1995). Probability judgment in medicine: Discounting unspecified probabilities. *Medical Decision Making, 15,* 227–230.

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.

The World Health Report: Working Together for Health. 2006. The World Health Organization (WHO).