Medical Problem Solving and Post-Problem Reflection in BioWorld

Sonia Faremo

McGill University

A thesis submitted to McGill University in
partial fulfillment of the requirements
of the degree of PhD (Doctorate) in Educational Psychology

0-494-06294-0

Your file Votre référence
ISBN: 0-596-00283-1
Our file Notre référence
ISBN: 0-596-00283-1

ABSTRACT

This study examined diagnostic problem solving and post-problem reflection in medical students, residents, and experts. Participants worked on three internal medicine cases from the computer-based learning environment, BioWorld. The analyses focused on general performance measures, problem solving operators and knowledge states, and post-problem reflection activities. Verbal protocol data was collected and examined using a coding scheme developed and implemented with the N-Vivo software. Students and residents differed in overall diagnostic accuracy, and significant differences were found in solution time and the number of utterances made for cases of varying difficulty. Differences in the use of operators and knowledge states are highlighted, although the groups were quite similar on many measures. The experts spent considerably more time working on case history information, consistently engaged in planning, and always generated the correct diagnosis (among others) in response to case history information. During post-problem reflection students used more case history data than residents. Expert models highlight the experts' problem solving cycle that consisted of reviewing data, identifying hypotheses, and planning. Post-questionnaire results indicate that participants found the cases to be interesting, useful for learning, but not especially difficult. Finally, several implications are drawn for the future development of BioWorld for medical training.

RÉSUMÉ

Cette étude explore la résolution de problèmes menant au diagnostique médical ainsi qu'à la réflexion entourant ce processus chez les étudiants en médecine, les résidents et les experts. Les participants ont travaillé à résoudre trois cas de patients en médecine interne présenté à travers l'environnement d'apprentissage par ordinateur BioWorld (CBLE ou computer-based learning environment). L'analyse des résultats se concentre sur l'évaluation générale des performances, les opérateurs aidant à la résolution de problème, les différents stades de connaissance et les retours et réflexions qui suivent la résolution de problème. Les protocoles verbaux furent analysés à l'aide d'un code d'analyse développé et implémenté avec Nvivo. Les résultats concernant les étudiants et les résidents diffèrent sur la fidélité et la validité de leur diagnostiques en général. Il y a aussi une différence significative de temps et du nombre d'énoncé émis pour des cas de niveau de difficulté variable. D'autres différences ont été soulignées par cette étude concernant l'utilisation d'opérateurs et des stades de connaissances entre les groupes, toutefois d'autres aspects de la résolution de problème se sont avérés similaires. Les experts dédient considérablement plus de temps à étudier les informations présentes dans l'histoire de cas, planifient systématiquement leurs activités, et génèrent entre autre invariablement le bon diagnostic en réponse à l'information présenté dans l'histoire de cas. Lors des réflexions suivant le diagnostic les étudiants utilise plus les données provenant de l'histoire de cas que les résidents. Le modèle expert bâtit met en évidence que le cycle de résolution de problèmes chez les experts consiste en une révision des données, l'identification des hypothèses, la planification et l'exécution de ces plans. Certaines données additionnelles recueilli lors du questionnaire suivant l'expérimentation indique que les participants on trouvé les cas intéressants et enrichissant mais pas nécessairement difficile. Plusieurs suggestions sont données pour le développement futur du système d'apprentissage informatisé BioWorld pour un environnement médical.

ACKNOWLEDGEMENTS

This study was only made possible with the assistance of a group of exceptional people. I am indebted to them all for their help in the various aspects of the study.

First, I am extremely grateful for the quality supervision provided to me by Dr. Susanne Lajoie throughout my doctoral studies. Her support and assistance has been outstanding throughout this period. From the time of early coursework through all the stages including a two-year absence I spent doing contract work, she has consistently been both supportive and helpful. Whenever I have submitted work to her for review, I have always received timely and detailed feedback, for which I am very appreciative. Further, her example of an academic who is an active researcher and administrator, and who really cares about students and the quality of academic work has been an inspiration to many graduate students. Also, I am very thankful to have worked on a study related to BioWorld, a system that has inspired me with its potential as a learning environment.

Second, many thanks also to Dr. Alenoush Saroyan for her support throughout my studies, and particularly for her rapid and detailed review of the dissertation manuscripts. Thank you for your many very helpful comments.

Third, in terms of conducting the present study, I owe a large debt of gratitude to several physicians affiliated with McGill's Faculty of Medicine. Dr. Jeffrey Wiseman was particularly instrumental in the study, having willingly devoted many hours as well as his medical expertise to assist in study design and data analysis. It would not have been possible to conduct the study without the assistance he provided. I look forward to future work that will build on the many discussions we had concerning computer-based learning environments for medical training. Dr. David Fleiszer and colleagues also supplied assistance in early design, recruiting participants, and analyzing the results of the study. Dr. Linda Snell also gave her time to participate in the study, and provided many useful comments concerning content and the learning activities in BioWorld.

Finally, Dr. Marguerite Roy was also instrumental in final coding decisions and in establishing inter-rater reliability. For careful review of early versions of the coding system, a thank you to Dr. Robert Bracewell. Thanks also to Andrew Chiarella for feedback on analyses, and Genevieve Gauthier for stimulating discussion and assistance with translation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1: OVERVIEW OF THE STUDY

Introduction

During their medical training, aspiring physicians are faced with the need to acquire tremendous amounts of complex knowledge, to develop and refine their problem solving and reasoning abilities, and to acquire a complex range of practical skills. In recent decades, medical educators have realized a need for innovations in medical training to assist students in learning this complex material. With the intention of providing better learning opportunities for students, many medical schools have adopted Problem-Based Learning (PBL) as their dominant instructional model (Barrows, 1996). In general terms, this approach emphasizes learning in context so that what is learned becomes situated in practice early, rather than in later internships or residency training. Students in PBL groups become actively engaged in solving patient problems in group learning situations. PBL has been an important instructional model for medical training for the past several decades.

The current study investigates cognitive processing in another type of "problem-based" learning environment for medical training. Essentially, it examines the processing that occurs as medical students and residents work on solving internal medicine cases from BioWorld (Lajoie, 2003), a prototype computer-based learning environment (CBLE). This system is designed to provide medical students with the opportunity to actively diagnose medical cases in a safe, supportive environment; this environment shares some similarities with other PBL environments but is designed primarily for individual problem solving. This study, therefore, is intended to identify how students and residents use case history, physical examination and diagnostic testing data in this particular problem-based environment to solve cases. It also examines how BioWorld might be modified to more effectively support learning in these populations.

PBL approaches are consistent with current educational and instructional models and principles. In Educational Psychology generally, the dominant paradigm has changed from *Behaviorism* (Skinner, 1957) during the middle portion of the century to *Cognitivism* (Clark, 1998; Cobb & Bowers, 1999; Mayer, 1992). More recently, *Constructivism* and principles of *Situated Learning* (Lave & Wenger, 1991) have widely been adopted (Clark, 1998; Cobb & Bowers, 1999). These models emphasize learners' active participation in authentic or realistic activities for making learning meaningful.

This chapter introduces the research study by providing a general description of BioWorld, a rationale for the study and its theoretical basis, the research methodology, the expected contributions of the study, and the research questions.

BioWorld: General Description

BioWorld is a prototype CBLE that has been developed during the last decade as an environment for learning high school Biology (see Lajoie, 2000, 2003). It is problem-based; students work with patient cases and attempt to solve them by finding the correct diagnosis. Solving a case involves working through a series of learning activities, consisting of problem solving (collecting and examining data), followed by post-problem reflection (reviewing the information they used during problem solving and comparing it with that of an expert). The system's general design is rooted in current cognitive and constructivist approaches to learning, in that it supports students as they actively engage in solving problems. See *Chapter 3* for a detailed description of BioWorld.

Previous BioWorld research was conducted in the domain of high school Biology (Lajoie, Lavigne, Munsie, & Wilkie, 1998). The current study examines BioWorld's first use with medical populations, undergraduate medical students and residents. This study examines the cognitive processing these participants engaged in, relates the findings to existing literature, and finally makes recommendations to aid the ongoing development of BioWorld for undergraduate medical education.

Rationale

This study examines medical problem solving embedded in a particular learning environment, BioWorld. While many studies of medical expertise have been conducted, few have employed realistic diagnostic problems as experimental tasks. Further, these studies have typically been conducted in order to investigate the components of expertise and not to improve learning in medical education. This study is a variation of the typical medical expertise study in that the experimental task is also a learning activity; the ultimate goal of the research is to support learning from engaging in this activity. The analyses and results therefore are two-fold, including both characterizations of cognitive processing and implications for learning environment design. For a variety of reasons, discussed in the following chapter, existing medical expertise research does not provide an adequate basis for learning environment design.

The results of this study add to the medical expertise literature, in part because the BioWorld case format elicits more realistic diagnostic problem solving as compared to the experimental tasks used in many studies. In BioWorld, problem solving proceeds in three phases that correspond to real-world diagnostic situations (case history, physical examination, and diagnostic testing). In addition, the diagnostic testing component allows learners to order test results as they would in actual practice. In real settings, a physician must determine what information to seek in order to solve a case, including case history, physical examination, and diagnostic test results. In contrast, other studies have often used cases consisting of textual descriptions of real patients delivered either all at once or sequentially in segments. The thinking involved in solving textual case descriptions logically differs in some respects from actual diagnostic situations. The features of the task are somewhat different when the important information has been pre-determined by the researcher(s) and presented to the participants.

Incorporating the results and recommendations from this study into BioWorld's design may make it an alternate, safe learning environment where medical students can effectively practice diagnosis and develop their knowledge and skills. Well-designed computer-based systems may provide possibilities for additional practice with realistic patient problems, as well as interactivity (e.g., students can interact with simulated patients, medical tools, etc.), and adaptivity to individual learning needs. Computers are uniquely well-suited for providing these important features, but these capabilities have not as yet been developed far enough in medical education.

Contributions of the Study

The results will contribute to our understanding of medical problem solving and of teaching and learning in medicine. Studying learning in complex, but constrained settings allow for closer examination of cognitive processing. Lajoie (2003) emphasizes the importance of understanding the developmental course of learning *or learning along a trajectory*. In this study, the differences between the groups are examined, and several implications for the development of BioWorld are identified. More generally, the results also have implications for research on medical problem solving and computer-based instruction for complex domains.

This study will extend previous medical expertise research by focusing on enhancing learning in one standardized learning environment. That is, while a significant body of cognitive research has been conducted examining a wide array of factors involved in medical problem solving, comparatively little work has been done on incorporating the results of these studies into medical training. Even less work has been done to develop computer-based instruction that supports diagnostic problem solving. In the best case scenario, the results will actually enhance the design of BioWorld for medical personnel and provide the basis for future empirically-based studies.

Theoretical Basis of the Study: Cognitive Science

The Cognitive Science Perspective

Overview

Cognitive science is a relatively new science, the beginnings of which can be traced to some important scientific advances in mathematics and computation, neuronal models, cybernetics, psychology, information theory, and neuropsychology that were made earlier this century (Miller, 1956; Wiener, 1948). The birth of cognitive science is generally marked by the 1956 meeting of several prominent researchers including Chomsky, Minsky, Miller, Newell, and Simon at MIT (Bruer, 1998; Newell, 1990). These researchers delivered a series of papers with a common theme, *the mind as a processor of information*, that countered the dominant behaviourist paradigm of the time (Bechtel, Abrahamson, & Graham, 1998).

According to the *computer metaphor* or the *computational view*, the mind is a physical symbol-processing system and human cognitive processes are analogous to the manipulation of data as performed by computers (Newell, Shaw, & Simon, 1958; Newell & Simon, 1972; Simon, 1990). *Cognitive Science* is defined by Simon and Kaplan (1989) as “the study of intelligence and its computational processes in humans (and animals), in computers, and in the abstract” (p.2).

The Cognitive Science paradigm has provided a basis for modeling and studying complex cognition for decades, and the findings from research based on this view are important both for theoretical development and for their philosophical implications (Lakoff & Johnson, 1999). It has been the dominant paradigm in related fields including psychology and educational psychology, and has provided a theoretical and methodological framework for studying problem solving and expertise in many domains (for a recent review, see Ericsson, 2003). More recent theoretical developments in education and related fields include *Constructivism*, and *Situated Learning* (Lave & Wenger, 1991) and some of these theorists regard Cognitive Science as incompatible with these views. However, a more moderate and probably more pervasive view is that cognitive theory needs to be broadened to take social and situational factors into account (Schoenfeld, 1999; Vosniadou, 1996). Some prominent researchers have indeed concluded that both perspectives are fundamentally important for education and should be pursued in educational research (Anderson, Greeno, Reder, & Simon, 2000).

This study merges these perspectives in a specific way. The problem solving that is examined is situated in a particular learning environment, BioWorld. The general design of the system is based on the belief that active participation in a domain is critical for learning to practice. The methods for examining problem solving in this study are drawn from Cognitive Science and Information Processing Theory (IPT); the analysis focuses on individual cognition, specifically how and what knowledge is worked through as individual participants solve BioWorld cases. A brief review of the main cognitive structures and processes specified by IPT is provided in this section. The results of the study may later be extended through further research and theoretical developments to involve learning from BioWorld in collaborative settings (e.g., Engestrom, 1995).

Information Processing Theory (IPT)

Newell and colleagues (Newell, et al. 1958; Newell & Simon, 1972) introduced IPT as a theory of problem solving that specified information and its processing as key elements of systems (e.g., people, computers) working on tasks (e.g., problem solving). The theory identified different types of memory, primitive information processes that work on information stored in memory, and a set of rules for combining processes into "programs" of behavior. According to IPT, cognition can be viewed in terms of *inputs* or stimuli (declarative knowledge), cognitive *processes* and *structures* involved in solving a problem, and *outputs* or actions (e.g., solutions) that result (Newell & Simon, 1972; Von Eckardt, 1993). Several fundamental elements of the theory are briefly reviewed in this section.

Physical symbol system hypothesis. Simon (1996, 1990) describes the *physical symbol system hypothesis*, which states that in order to be capable of thinking, a system must be able to perform basic functions using symbols (input, output, store, construct, modify, erase, compare, and select operations on the basis of comparison). The major focus is the explication of these states and the processes involved in the transition between them (Massaro & Cowan, 1993). The fundamental unit learned and stored in human memory is believed to be the semantic unit, which is meaningful and is stored in the form of propositions.

Architecture. The cognitive or *symbolic rule-based architecture* is a central construct within cognitive science (Anderson, 1991; Newell, 1992, 1993; Newell, Rosenbloom, & Laird, 1989; Pylyshyn, 1996). A *cognitive architecture* is the set of fixed structures or the framework that supports human knowledge processing (Laird & Rosenbloom, 1996). The human cognitive architecture supports five general functions: memory, symbols, operations, interpretation, and interaction with the external world (Newell et al., 1989). The following three levels are assumed for the cognitive architecture and for computer models of it: the knowledge or semantic level, the symbolic level, and the physical or biological level (Pylyshyn, 1989). Computer models of the human cognitive architecture have been a major area of cognitive science research (Anderson, 1983, 1993a, 1993b; Newell & Simon, 1972; Newell, 1990). The majority of expertise research has focused on knowledge level representation.

Memory. Another key construct for IPT is memory. Memory structures support key IPT processes including encoding, storage, and retrieval (Best, 1995). Broadbent's (1958) model specifies *long-term memory* (LTM) as the permanent storage, and *short-term memory* (STM) as the location or temporary storage for information currently being processed. The *sensory register* is the brief storage for auditory and visual stimuli. The influential Atkinson and Shiffrin (1968) model specifies these fixed structures and the flexible structures (e.g., control processes) operating on them. Craik and Lockhart's

(1972) levels of processing theory according to which semantic processing results in deeper and more permanent memory has also been influential in memory research.

Representation. A related construct is knowledge or *mental representation* (Rumelhart & Norman, 1988; Suppes, Pavel, & Falmagne, 1994). Mental representations contain information and interpretations concerning objects and events from the external world, they exist as part of a representational system, and rules map the external world onto the represented world (Markman, 1999; McNamara, 1994). A common distinction between types of knowledge has been *declarative* versus *procedural* (Anderson, 1993b), although it is regarded as controversial (Mandler, 1998). McNamara (1994) provides a taxonomy of mental representations based on these two categories. First, it specifies two classes of declarative knowledge: 1) *analogical* or perceptual representations (visual, spatial, etc.), and 2) *symbolic* or propositional representations. Second, procedural representations consist of *productions.* Several forms of representation have been studied in cognitive research and in medical expertise research. Included are scripts, schemas, mental models, and perceptual representations (Bartlett, 1932; Johnson-Laird, 1989; Kosslyn, 1996; Schank & Abelson, 1977).

Production systems. Another key construct for cognitive science is the *production system*, which consists of a set of If...Then rules (condition-action pairs). These rules form the basis of symbolic computational models (Newell, 1990; Schunn & Klahr, 1998). When its associated conditions are satisfied, an action will fire. Production systems have been the basic mechanism for modelling problem solving and other psychological processes (Hunt, 1989). Two data structures or memory stores, LTM and STM, and three processes (recognition, conflict resolution, and action) support production systems (Klahr & MacWhinney, 1996).

Two well-known examples of cognitive architectures that have been implemented as computer-based production systems are ACT-R (Anderson, 1983, 1993b; Anderson & Labriere, 1998) and Soar (Laird & Rosenbloom, 1996; Laird, Newell, & Rosenbloom, 1987; Luger, 1994; Newell, 1990; Newell et al., 1989). These systems and others (e.g., EPAM, the General Problem Solver or GPS) are based on IPT and share many similarities (Newell et al., 1989; Simon, 1996).

Goals and plans. According to IPT, behavior is goal-based. Given a current or actual state in a task environment, a *goal* specifies another state that is desirable and to be achieved. In a classic paper, Miller, Galantner, and Pribram (1960) defined *planning* as a structured organization in behavior, and a *plan* as a "hierarchical process in the organism that can control the order in which a sequence of operations is to be performed" (p. 178). Generally, a plan has been defined as a set of goals that together specify the means to achieve a desired end state (Bracewell & Breuleux, 1994).

There have been varied definitions and three prevalent descriptions of planning in the cognitive psychology literature: a general model for the control of problem solving, a component of problem solving, and a schema for understanding goal-directed behavior (Friedman, Scholnick, & Cocking, 1987). Friedman et al. conclude that, despite its importance as a "central executive function" (p.533), planning has been somewhat elusive as a topic of scientific investigation. Research has established that experts often spend more time planning than executing tasks, but research beyond that has been limited. Finally, they also note that many factors (social, affective, and cognitive) influence the way people plan in different domains.

Various models of planning have since been developed, e.g., the *Opportunistic Model of Planning* (Hayes-Roth & Hayes-Roth (1979) specifies planning occurring in two stages during problem solving: 1) planning and 2) monitoring plan execution. It also specifies planning as *opportunistic* (plans can be changed in light of new data) and *multidirectional* (proceeding through different levels of abstraction). Finally, five levels of abstraction are specified in the model: plan, plan abstraction, knowledge base, executive, and metaplan.

Problem Solving

IPT approaches have been applied to investigate a broad range of cognitive phenomena (Best, 1995; Neisser, 1967, 1992). A major one has been *problem solving*, which can be defined as the cognitive processing involved in achieving a goal when no solution is immediately apparent (Dunbar, 1998; Ericsson & Simon, 1993). Problem solving is the process of interest for this study, and it is a major focus of cognitive science research.

Problem space. A key construct in cognitive science, the *problem space* (Newell & Simon, 1972), is made up of an initial state, a goal state, and a set of operators that the problem solver can apply. Solving a problem then is viewed as a process of working through a series of problem states or representations from the initial state to the goal state.

Heuristics and strategies. Heuristics are problem solving "rules of thumb" that can be applied during problem solving to advance closer to the goal state. A variety of heuristics have been studied; examples are *hill-climbing* and *means-ends analysis* (see Stanovich, 2003). Cognitive strategies or reasoning have also been widely examined. *Forward* or *top-down* (data-driven) reasoning refers to a process of working from data to a solution. It is usually considered a strong method and is associated with expert performance. *Backward or bottom up* (hypothesis-driven) reasoning refers to working from hypotheses to a solution; it is generally a weak method used by non-experts.

Problem types. Important distinctions between domains and problem types include *ill-structured* versus *well-structured*, and *knowledge-rich* versus *knowledge-lean* (Kotovsky, 2003; Lesgold, 1988; Simon, 1973; VanLehn, 1989; Voss & Post, 1989). Well-structured problems have a single solution and established procedures for achieving them (e.g., mathematical equations), while ill-structured tasks (e.g., writing) do not. Knowledge-rich tasks require domain-specific knowledge that cannot be derived from the problem (e.g., medical diagnosis), while knowledge-lean tasks (e.g., cryptarithmetic) do not. In the case of medicine, diagnosis has been regarded as involving knowledge-rich, ill-structured problems (one unspecified solution and many possible paths to reach it).

Summary

Many studies have worked on eliciting the characteristics of these basic processes and structures in medicine and other domains. This brief review provides a basis for understanding the methods and results of this study. In addition, several of the structures and processes discussed in this section are also identified in *Chapter 2.*

Research Methodology

The "Think-Aloud" Method and Verbal Protocol Analysis

In addition to the theoretical framework, cognitive scientists have developed a number of methodologies for studying human cognition (see Lakoff & Johnson, 1999). A standard approach for studying problem solving is *verbal protocol analysis* or the standard think-aloud approach (Bracewell & Breuleux, 1994; Chi, 1997; Ericsson, 1998; Ericsson & Simon, 1993) which has been used in many studies of expertise in medicine and other domains (e.g., Hassebrock & Prietula, 1992). Briefly, it involves having participants attempt to solve a problem while thinking out loud. Their verbalizations are recorded and later analyzed. Coding systems are developed to allow researchers to systematically examine the content of the verbalizations that occurred during problem solving.

A set of basic assumptions underlying this methodology are incorporated into the analysis and the coding scheme used in the study.

1. Problem solving behaviour is a search through a problem space.
2. Each step in the problem solving process is accomplished through application of an operator.
3. The contents of verbalizations corresponds to some part of the contents of short-term memory (STM).
4. The information in STM consists of knowledge (inputs), new knowledge produced by the application of operators, and symbols representing active goals and sub-goals.

In this study, the two main coding categories are operators and knowledge states. The knowledge states further break down into 1) directly observable data and 2) inferred knowledge that results from an operation on data or is recalled from long-term memory. This will be described in more detail in *Chapter 3: Method.*

The Domain: Medical Diagnosis

Over the course of their training, medical students acquire a large body of complex knowledge and diverse skills, and they learn to apply their knowledge to diagnose a wide range of medical case types. They learn to perform *differential diagnosis*, a process of working through data to a solution, acquiring pertinent additional information when required, and coming up with a diagnosis that may include several possibilities or hypotheses (Cimino, 1999; Kassirer, 1995; Kassirer & Kopelman, 1991; Kuipers & Kassirer, 1984; Patel, Arocha, & Kaufman, 1999b). They learn to carry out this same general process, using it in a number of settings and with a wide range of patient problems.

At a general level, medical diagnosis is similar to diagnostic procedures in other domains such as electronics troubleshooting (Lesgold & Lajoie, 1991). That is, a diagnostician in either domain is faced with a "malfunction" of a complex system, and initially incomplete data, and s/he must rely on their domain knowledge to gather further data and determine the nature of the problem. It differs from other domains in which problem solving research has been conducted where all of the elements needed for solution are given at the start (Elstein & Rabinowitz, 1993; Evans, 1989; Gilhooly, 1990).

Medical diagnosis has been construed as a problem solving activity (Elstein & Rabinowitz, 1993). At the general level, diagnosis is ill-structured since there is only one correct diagnosis for each patient problem. However, there are numerous paths to follow to reach that one diagnosis. IPT provides the means for conceptualizing medical diagnosis as problem solving and for modeling the cognitive processes and structures involved. Physicians will use several forms of processing including inference (Patel & Ramoni, 1997), as well as knowledge structures at varying levels of abstraction (Evans & Gadd, 1989). In addition, rather than being a unitary activity, there are variations in how diagnostic problem solving is carried out that relate to different sub-domains of medicine (e.g., dermatology, endocrinology), level of expertise, and problem type.

Research Questions

The major purpose of this study is to investigate diagnostic problem solving in BioWorld, and to draw implications for the future development of the system. To this end, the following specific research questions are addressed.

1. How does the general performance (time on task, protocol length, diagnostic accuracy) of students and residents differ as they attempt to diagnose internal medicine cases based on case history, physical examination, and diagnostic testing activities? How does their performance compare with that of experts?
2. How does the problem solving of students and residents differ as they attempt to diagnose internal medicine cases based on case history, physical examination, and diagnostic testing activities? How does their problem solving compare with that of experts?
3. How does the post-problem reflection (categorizing data, prioritizing data, and comparing to an expert list) of students and residents differ? How do their reflective actions compare with those of experts?

Implications of the Study

This research study focuses primarily on characterizing cognitive processing during problem solving and post-problem reflection. In addition, the results of the study will be used to draw instructional implications for the ongoing development of BioWorld as a learning environment. *Chapter 5* identifies some of the instructional implications of the current findings, as well as existing research that may be relevant to BioWorld's future development. The results will support the ongoing design of learning activities in general, as well as the design of more specific tools and functionalities.

Chapter Summary

This introductory chapter has presented an overview of several aspects of this study. A basic overview and the rationale for both the larger development effort and the current study were provided. Also included were a basic description of Cognitive Science as the theoretical framework and the use of its research methods were reviewed. Medical diagnosis was briefly characterized as a complex, knowledge-rich problem solving task. The following chapter reviews cognitive research on medical expertise, some of the implications for medical education drawn from medical expertise research, and research on computer-based instruction in medicine.

CHAPTER 2: LITERATURE REVIEW

Introduction

This review briefly summarizes general research on problem solving and expertise and focuses primarily on the findings from several decades of medical expertise research. In order to provide a context for this study, the review also contains a brief overview of research on technology in education, focusing on applications for medical education.

Problem Solving and Expertise Research

For several decades, cognitive scientists have examined the cognitive processing involved in solving complex problems. The literature now contains a large body of research on the problem solving of novices and experts (see Alexander, 2003; Davidson & Sternberg, 2003; Wagner, 1999). Research results include various characteristics that distinguish experts from novices, which have repeatedly been summarized (see Ericsson, 2003; Ericsson & Charness, 1994; Gilhooly, 1990; Glaser & Chi, 1988). Not surprisingly, findings confirm that domain-specific knowledge plays a critical role in expert performance, and expertise develops through deliberate practice.

Research results also suggest that the *quality of internal problem representation* is a critical characteristic of expertise. Expert knowledge is apparently highly structured and organized in a larger number of schemas and more efficient chunks than the knowledge of non-experts. This knowledge is also conditionalized or "tuned" to situations. Experts have been found to interpret and depict problems at a deep level and to be able to recognize meaningful patterns. Other established characteristics of experts as compared to non-experts include *superior memory recall*, and *fast and highly effective performance*. Expert knowledge is also often *partially procedural, implicit or tacit*. In addition, experts carry out *qualitative analysis of problems prior to solving them,* and they demonstrate *superior self-monitoring and self-knowledge skills*. A more detailed review of expertise research in the domain of medicine is provided in the following section.

Overview of Medical Expertise Research

As for expertise research in general, two main research approaches have been applied to the study of medical cognition: *problem solving* (see Lakoff & Johnson, 1999; Newell & Simon, 1972; Patel, Kaufman, & Arocha, 2002) and *decision making* (see

Elstein, 2001, 2000; Elstein & Schwarz, 2002; Falzer, 2004; Lipshitz, Klein, Orasanu, & Salas, 2001; Patel et al., 2002). The current study follows the problem solving paradigm, as described in Chapter 1. The role of a number of cognitive processes (e.g., hypothesis generation) and structures (e.g., scripts, schemas) at different levels of abstraction have been investigated, including general competencies applied during problem solving (e.g., comprehension), as well as specific diagnostic competencies that develop over a longer term (e.g., knowledge encapsulation). Various general measures such as solution time and diagnostic accuracy have also been investigated.

Methodology in Medical Expertise Research

Medical expertise research now spans roughly five decades. Two general eras in medical expertise research correspond to two dominant paradigms in Psychology during this time frame (see Patel & Arocha, 1995). Early studies, following the *Behaviorist* paradigm, largely assessed clinical skills using psychometric tests. With the *Cognitive* paradigm, the focus of medical expertise shifted towards examining problem solving. Two general methodologies have frequently been used in these studies, *recall* and *think-aloud.* Variations on these basic experimental tasks have frequently been used, and other experimental tasks have also been used. Several methodological concerns will be addressed later in this review.

Recall Studies

A number of researchers have investigated medical thinking by using the standard *free recall* methodology borrowed from other psychological studies (see Vincente & Wang, 1998). This approach was adopted fairly early in medical expertise research and maintained by several prominent researchers for many years. It involves presenting a text (e.g., written case description) to a participant for a specified period of time, removing it, and then asking the participant to recall as much as possible about the contents of the text. The information recalled is subsequently analyzed and compared to the original text, and the contents and/or structure of memory are examined. This methodology and variations of it have been used in numerous studies, such as the *explanation protocol* task, whereby participants are asked to recall a case, diagnose it, and to explain underlying pathophysiology (Patel & Groen, 1986; Schmidt & Boshuizen, 1993).

Think-Aloud Studies: Verbal Protocol Analysis

The standard expertise approach involves *examining cognitive processing as it takes place*; participants with different levels of expertise (e.g., experts, novices, and often intermediates) individually solve problems while thinking out loud. In the usual scenario, also described by Cuthbert, du Boulay, Teather, Teather, Sharples, and du Boulay (1999), participants are individually presented with a set of cases, one after another, in print form or on a computer screen. Case information may be presented in text form all at once, or in sequential segments. Participants are asked to "solve" the cases while thinking aloud, and to provide a diagnosis for each case. Verbalizations during problem solving are recorded. Time limits for presentation have often been imposed and, in some cases, probes have been used to elicit more verbalization at specific points during problem solving. Typically, verbal protocol analysis methods (Ericsson & Simon, 1993) have been used to analyze transcribed data. This involves the development and application of a coding scheme that enables systematic analysis and comparison of the thinking of individuals with different levels of expertise. Other analytic techniques have also been employed (e.g., propositional analysis, semantic networks).

Medical Expertise Research Findings

This review samples the findings from a variety of studies conducted over approximately the past three decades, and provides a sense of the varied research efforts that comprise this field. It is organized in four sections that cover the following research: 1) general measures such as diagnostic accuracy, reading time, solution time, 2) general problem solving strategies (e.g., hypothesis generation), and 3) memory and knowledge structures and knowledge types.

General Measures

Several general measures, including diagnostic accuracy, case solution time, and case reading time have been addressed in a number of studies.

Diagnostic Accuracy

Diagnostic accuracy in medical problem solving has been among the dependent variables in most medical expertise studies. Experts are more accurate in diagnosing cases than non-experts (e.g., Patel & Groen, 1991b), and a linear increase with expertise has often been established (e.g., Norman, Trott, Brooks, & Smith, 1994). Some interesting exceptions are summarized in this section.

In studies that include intermediate-level participants, some researchers have found a U-shaped relationship between diagnostic accuracy (and other measures) and expertise. That is, instead of the expected linear increase, in some cases the diagnoses of novices and experts are more accurate than those of intermediate level participants. In their classic study, Lesgold, Rubinson, Feltovich, Glaser, Klopfer, and Wang (1988) reported that 1^{st}-year residents performed almost as well as experts solving typical ("textbook") radiology cases. Senior residents apparently knew that other potential diagnoses needed to be considered, but were not knowledgeable enough to accurately choose between the possibilities. Novices, on the other hand, used their basic knowledge of standard or typical disease presentations to diagnose more accurately.

Another set of relevant findings shows that experts in one sub-domain of medicine do not perform at an expert level when attempting to diagnose cases from another sub-domain. For example, Rikers, Te Winkel, Loyens, and Schmidt (2003) studied medical students, pulmonologists (sub-experts), and cardiologists (experts) solving cardiology problems. They found a statistically significant linear increase in expertise from students to sub-experts and from sub-experts to experts. Other findings concerning accuracy are addressed during the course of this review.

Case Solution Time

As expected, experts have repeatedly been found to perform faster than novices on routine cases. For example, in a study with 6^{th}-year students and experienced family physicians, Custers (1995) found a main effect for case type and reading time with level of expertise. In addition, some studies have shown that solution time is shorter for typical versus atypical cases (e.g., Custers, Boshuizen, & Schmidt, 1996).

Other findings on solution time include the inverted-U relationship between increasing level of expertise and time required to solve a case. That is, novices and experts have required less time than intermediate participants to solve cases (Lesgold et al., 1988; Norman, Brooks, & Allen, 1989; Rikers, Schmidt, & Boshuizen, 2000). Schmidt and Boshuizen (1993) report that intermediates have more difficulty with time constraints than novices or experts. As for diagnostic accuracy, it is possible that these findings may reflect difficulty on the part of intermediates in selecting between possible explanations and diagnoses. It should be noted that the experimental tasks used in these various studies have differed in ways that could differentially impact solution time.

Case Reading Time

Other studies have examined case reading time only (time to read through case information), and found that experts were faster than non-experts (Claessen & Boshuizen, 1985; Rikers, et al., 2000). For example, Custers (1995) found that 6th-year medical students consistently took longer (14% longer overall) regardless of case typicality to read a case description than did experienced family physicians. In another study, cardiology problems were divided into medical history, physical examination, laboratory data, and additional findings to examine the role of each in diagnosis (Rikers, et al., 2003). Two less-experienced groups (students and nephrologists) needed considerably more time to process the case information than the experts. However, Norman, et al. (1994) studied 1st-year residents in family medicine, 1st- and 2nd-year residents in internal medicine, and experts solving difficult nephrology problems. In their study, the experts took more time to review the cases than the other groups, which the researchers attribute to their more elaborate (think aloud) explanations.

Problem Solving Approaches

Early efforts in medical cognition research focused on the identification of general strategies or skills as determinants of medical diagnostic ability.

The Hypothetico-Deductive Approach

An early and very influential model of diagnostic problem solving was the *Hypothetico-Deductive* method, presented by Elstein, Shulman, and Sprafka (1978). According to this approach, hypothesis generation and testing are key determinants of successful medical diagnosis (Elstein, Shulman, & Sprafka, 1990; Schwarz & Griffin, 1986), and the diagnostic process involves 4 activities: *cue acquisition, hypothesis generation, cue interpretation*, and *hypothesis evaluation*. Using this method, a physician initially generates hypotheses based on case information, considers which symptoms should be present for each of the hypothesized diseases/conditions, and then checks to see whether the symptoms are present for each.

Subsequent research established that the hypothetico-deductive approach may not result in effective performance (Claessen & Boshuizen, 1985; Gale & Marsden, 1983; Neufeld, Norman, Feightner, & Barrows, 1981; Schwarz & Griffin, 1986). Researchers have suggested that use of the this method may interfere with learning to reason in a forward manner (Patel et al., 1999) and is actually not an effective diagnostic strategy for complex cases where patients have multiple problems (Custers et al., 2000). The emphasis on hypotheses rather than data is seen as a fundamental flaw. Subsequent research indicates that the use of data rather than hypotheses is associated with expert performance (Patel, Arocha, & Kaufman, 1994). In later work, Elstein (1994) presents a hybrid model of reasoning, and suggests that several factors affect how a clinician solves a problem including the perceived difficulty and characteristics of the problem and skill level. For example, simple problems may be solved by pattern recognition, while more difficult problems may require explicit hypothesis testing.

Gale and Marsden (1983) proposed an alternative to the hypothetico-deductive method. They accounted for diagnostic thinking in terms of two cognitive processes, *structuring* and *extrapolation.* According to this view, a clinician actively imposes order or structure on the clinical data by extrapolating to prior knowledge. Structuring is seen as a reciprocal, dynamic process involving many modifications over time. Structured knowledge includes both content and procedural components. Extrapolation is seen in the interpretation of signs and symptoms and their importance for diagnostic performance. These researchers also propose that the diagnostic thinking process proceeds through three stages: *initiation of interpretations, progress of diagnostic thinking processes*, and *resolution.* Each stage has specified thinking processes (e.g., interpretation, inquiry) and psychological factors (e.g., cognitive restructuring).

Hypothesis Generation and Evaluation

An established characteristic of medical expertise is the tendency to generate the correct hypothesis among the differential diagnoses from very early in the problem solving process. Less experienced individuals do not reliably do so. Some participants fail to generate the correct diagnosis, while others fail to correctly distinguish between the correct hypothesis and other competing hypotheses (Johnson, Duran, Hassebrock, Moller, Prietula, Feltovich, & Swanson, 1981). Neufeld, et al. (1981) concluded that if experts consider the correct diagnosis within 5 minutes, they become definite about the diagnosis in 95% of cases; however, if the correct diagnosis is not considered within 5 minutes, there is a 95% probability that it will be missed.

Other studies provide more detail concerning how hypothesis generation occurs at different levels of expertise (Arocha & Patel 1995b; Joseph & Patel, 1990; Patel & Groen, 1986). Patel et al. (1994) found that three general strategies might be associated with diagnostic problem-solving. They were: 1) coming up with a major hypothesis and subsequently reinterpreting it or ignoring contradictory information, 2) generating hypotheses concurrently to explain different data, and 3) developing multiple initial hypotheses and subsequently narrowing the set of possible ones. Experts developed hypotheses quickly and accommodated subsequent data. In another study of novices at various levels in training, 2nd year students maintained hypotheses despite contradictory evidence, 3rd year students developed different hypotheses to account for different findings, and 4th year students initially generated several hypotheses, subsequently narrowing them down to a single coherent explanation (Arocha, Patel, & Patel, 1993).

Other studies have examined the role of domain knowledge in hypothesis generation, and obtained similar results. Joseph and Patel (1990) had endocrinologists (high-domain knowledge or HDK) and cardiologists (low domain knowledge or LDK) solve an endocrinology case. HDK participants generated links to organize the text, while the others did not. The groups gave similar diagnoses. After identifying the correct diagnosis, the HDK participants generated few additional hypotheses, ruled out some hypotheses, confirmed the diagnosis, and identified secondary problems. The LDK group continued to produce hypotheses, focused on test results, related them to new hypotheses, and failed to rule out hypotheses despite contrary evidence. In a study of, low-, medium- (MDK), and high-domain knowledge individuals, Patel, HoPingKong, and Mark (1984) found that the HDK participants' knowledge was more "specific" and was used more efficiently than that of the other groups. MDK participants had difficulty sorting relevant versus irrelevant information. LDK participants did not have the relevant knowledge and were not able to make useful inferences.

Medical students have been found to develop *more* initial hypotheses, while experts' hypotheses were *more general* (Sisson, Donnelly, Hess, & Woolliscroft, 1991).

Lesgold et al. (1988) found that novices appear to restrict hypotheses to the *most obvious*, but why they do so was not established. In another study, "super-experts" generated *more pertinent* hypotheses and maintained deliberate reasoning with atypical cases, both of which non-experts did not do (Raufauste, Eyrolle, & Marine, 1998); faced with atypical cases, the problem solving of non-experts became unconstrained.

Several studies have looked at the generation of hypotheses in response to clinical data. Johnson et al. (1981) looked at the numbers and types of hypotheses that were generated in response to history, physical examination, x-ray, and EKG data. They found that more hypotheses were generated in response to history and physical examination data than to x-ray or EKG data. Gruppen, Woolliscroft, and Wolf (1988), found that physicians (1st, 2nd, and 3rd-year house officers), generated the correct diagnosis based on chief complaint and personal information in over 90% of the cases. The researchers found that the chief complaint had a "generative" role, producing hypotheses that influenced subsequent data collection and evaluation. Focusing on medical students, Gruppen, Palchik, Wolf, Laing, Oh, and Davis (1993) most participants entertained the correct diagnosis as a possibility after learning the patient's current complaint and reading the patient description. Students who did not identify the correct diagnosis after reviewing the history were significantly less likely to correctly diagnose the case.

Problem Solving Strategies

An established finding in general and medical expertise research is that experts use *data-driven* (forward reasoning) strategies, while novices tend to use a *hypothesis-driven* (backward reasoning) or a *mixed strategy* (Barr & Feigenbaum, 1981; Lesgold, 1988; Norman et al., 1994; Patel, Groen, & Arocha, 1991; Patel & Ramoni, 1997; Reimann & Chi, 1989). Data-driven problem solving is considered a strong method that moves from data to a solution. In contrast, hypothesis-driven problem solving is a weak method whereby an individual moves from hypotheses to the facts concerning a case.

Other studies have shown that 1) intermediate-level individuals may use a mixed strategy, 2) diagnostic accuracy is associated with data-driven problem solving, 3) experts exhibit mixed-strategy use and make incorrect diagnoses outside of their specialty areas, and 4) experts use weak methods (backward reasoning) when their knowledge base is inadequate (Joseph & Patel, 1990; Norman et al., 1994; Patel & Groen, 1991; Patel et al., 1990). For example, Patel and Groen (1986) examined verbal protocol data from cardiologists and endocrinologists. They found that experts who made correct diagnoses used data-driven (forward) reasoning, while those who were incorrect used some hypothesis-driven (backward) reasoning. Similarly, Patel, Groen, and Arocha (1990) found that forward reasoning does not always accompany diagnostic accuracy, but that inaccurate diagnoses seem to always be accompanied by backward reasoning. Finally, the forward reasoning of experts breaks down and is replaced by backward reasoning when unrelated facts or "loose ends" are introduced and have to be explained (Patel, Groen, Ramoni, & Kaufman, 1992).

Strategy use by intermediates and experts solving difficult problems may involve more mixed strategy use. Radiology residents and expert radiologists were studied as they diagnosed difficult cases from case histories and mammograms (Azevedo, 1997). These groups predominantly used data-driven or mixed strategies depending on case typicality and level of experience. In the other study, Norman et al. (1994) found that all participants (1st and 2nd–year internal medicine residents and experienced nephrologists) used mixed strategies.

Types of Reasoning

Causal reasoning involves thinking about the possible pathophysiological causes underlying diseases (Kuipers & Kassirer, 1984; Patel & Groen, 1986; Patel, Evans, & Groen, 1989). While novices often make use of causal reasoning, experts make little use of it in explaining their diagnoses (Patel & Ramoni, 1997; Schmidt & Norman, 1990). Considerably less attention has been given to more general or "generic" forms of reasoning in medicine (such as *reasoning to alleviate to uncertainty*). Recently, researchers have distinguished between the use of 1) deliberate causal reasoning (explicit) or the "science" of medical practice, and 2) tacit knowledge (implicit) or the "art" of medical practice (Patel, Arocha, & Kaufman, 1999a). They suggest that implicit knowledge, which is important for expert performance, develops with experience.

Scheme-inductive reasoning has recently been proposed; a *scheme* (distinct from *schema*) is an organized structure for diagnostic reasoning (Coderre, et al., 2003; Mandin, Jones, Woloschuk, & Harasym, 1997). Briefly, it is a decision process that proceeds from a set of alternative causal groups to tests (at branching points), and then to the exclusion of some alternatives and acceptance of the remaining. For "learning and problem solving (schemes) provide the advantage of combining the creation of a knowledge structure and a search-and-retrieval strategy" (Mandin et al., 1997, p.173). Once the process has proceeded through multiple tests (evaluations of data), the number of diagnostic possibilities is reduced so that other strategies, namely deductive reasoning or pattern recognition can be used to complete the process. Coderre, et al. examined diagnostic "success" in relation to the expertise level, type of reasoning strategy employed, and clinical presentation in experts and non-experts. The reasoning strategy used was significantly related to diagnostic success. Pattern recognition was most highly associated with diagnostic success, followed by scheme-inductive reasoning. Participants who used hypothetico-deductive reasoning were much less successful. However, commenting on the Coderre et al. study, Norman and Eva (2003) suggest that experts more likely use a combination of strategies, rather than consistently using one strategy.

Another distinction is *depth-first* versus *breadth-first.* Depth-first refers to the process whereby a diagnostician maintains a single diagnosis, until disconfirming data are obtained. At that point, more hypotheses are generated. Breadth-first refers to the strategy of generating all important diagnostic possibilities early during problem solving. The performance of participants (students, residents, and experts) in the Johnson et al. (1981) study in pediatric cardiology indicated that neither strategy was more strongly associated with diagnostic success. Rather, the researchers claim that the ability to generate all possible alternatives and how patient data was "integrated" in relation to diagnostic possibilities were important for correctly solving a case.

Deep versus *shallow* reasoning has also been investigated (Patel et al., 1992). Deep reasoning refers to reasoning with causal representations of the internal operation of a system. Used by experts, it gives the problem solver explanatory and predictive possibilities (Patel, et al., 1994). Shallow reasoning refers to reasoning with surface-level aspects of a problem. While Patel and Ramoni (1997) note that there is little direct proof of "deep reasoning" in medical expertise research, experts do tend to categorize problems using deeper domain principles while novices categorize based on more surface features.

The *Select and Test Model* (STM) breaks medical diagnosis down into two phases 1) hypothesis selection, and 2) hypothesis testing (Magnani, 1992; Ramoni, Stefanelli, Magnani, & Barosi, 1992; Patel & Ramoni, 1997). It also identifies four types of inferencing that are involved in these phases: 1) *abstraction* to identify problem features, 2) *abduction* to select potential hypotheses based on problem features, 3) *deduction* to predict what should be true about the case for each potential hypothesis, and finally 4) *induction* to match predictions with the real state of a patient and find the best match.

Categorization or Pattern Matching

Medical diagnosis has also been described as a categorization or pattern matching task (Bordage & Zacks, 1984; Brooks, Norman, & Allen, 1991). From this viewpoint, expertise is dependent on having a large number of prior examples stored in memory that can be easily and rapidly retrieved in response to a case (Schmidt, Norman, & Boshuizen, 1990). Other researchers believe that viewing diagnosis as pattern recognition trivializes the complexity that is actually involved (Barrows & Feltovich, 1987). Pattern matching may play a more substantial role in sub-domains of medicine that have a significant visual component (e.g., dermatology). The topics covered in this review suggest that other forms of processing are important in medical problem solving.

Strategic Thinking

Focusing on relevant work in the domain of medicine, Evans and Gadd (1989) describe planning as important in problem solving contexts because it "establishes contexts for the resolution of uncertainty and provides a basis for the inferences required to interpret new information coherently" (p.231). They further identify three classes of goals that are used in diagnostic planning: 1) presenting items of case information in an accepted order, 2) coming up with hypotheses and support for them, and 3) establishing a case explanation that accounts for all of the evidence. In the Lesgold et al. (1988) study, expert radiologists diagnosing from chest x-rays quickly started to execute general plans. They looked at a film and quickly mentioned a diagnostic category (interpreted as the initiation of a general plan), and would then proceed to discuss the film mentioning relevant findings. Azevedo (1997) found that residents and expert radiologists were similar in their use of diagnostic planning (specifying further investigations), but experts used more goal statements (intended actions) than the residents. Azevedo's coding scheme was also applied to data from another radiology study involving medical students and residents, who were found to seldom engage in planning as compared to the experts in the original study (Azevedo & Faremo, 1999). In sum, experts may have more use for or facility with planning than non-experts.

Clinical Problem Analysis (CPA) characterizes medical diagnosis as a series of five steps: 1) obtaining relevant information about a patient from various sources, 2) constructing a list of "activating findings", 3) building a list of problems, 4) specifying a list of differential diagnoses, and 5) developing an action plan (Custers, Stuyt, & De Vries Robbe, 2000). CPA is presented as a teaching model that can provide a structure for students to learn and use in solving clinical problems. Unlike various other models, it presents a general set of "cognitive" steps to follow that can structure diagnostic problem solving. It also specifies a planning step (#5 developing an action plan), indicating that this is important for diagnostic problem solving in medicine. CPA will be discussed further in later chapters.

Finally, planning has been found to be important for medical tasks other than diagnosis. For example, Xiao, Milgram, and Doyle (1997) examined the preparatory planning of experts in Anesthesiology, noting that the experts' mental preparation seemed fragmented, and occurred in various forms (e.g., general decisions, concerns, overall action plans, specific actions). The researchers also recommend further study of planning is this domain.

Use of Clinical Case Information

Several studies have examined the use of clinical information (e.g., age, sex, profession, previous diseases) and have found that experts use it effectively and in an accurate manner, while novices are apparently unable to do the same (Custers, 1993; Custers, Boshuizen, & Schmidt, 1992; 1996; Hobus, Boshuizen, & Schmidt, 1990).

Other studies have focused on data-gathering in clinical interviews. Students, residents, and experts collected information during the clinical interview and used it to make a diagnosis (Kaufman & Patel, 1988). They solved one case of an individual with two medical problems. The experts conducted shorter interviews with the patient, elicited fewer findings and observations, but elicited approximately the same number of relevant findings as residents. Students elicited fewer relevant findings. Most experts diagnosed both problems, while no students did. Residents performed more variably: some recognized one problem only, while others recognized aspects of both. Patel, Groen, and

Patel (1997) examined how students, house-staff (intermediate), and experts conducted a patient work-up. Experts formed richly integrated knowledge structures based on the history taking and used them throughout problem solving. Less-integrated knowledge was found in housestaff, while students' knowledge only superficially resembled that of experts. Finally, "clinical interactions" between students, residents, and experts were studied by Allen, Arocha, and Patel (1998). The use of evidence was related to the early development of accurate hypotheses, and level of expertise does result in the use of different strategies for dealing with inconsistent evidence. Experts used both efficient evidence-gathering strategies and controlled reviewing of the evidence that resulted in refinements to initial hypotheses. Residents used predictive strategies also, but did not use evidence to refine or replace inaccurate hypotheses. Students used inefficient evidence-gathering, did not develop accurate hypotheses, and lacked knowledge of underlying pathophysiology.

Memory and Knowledge Structures, and Knowledge Types

A number of studies have sought to replicate the classic psychological studies that demonstrated a linear relationship between memory recall of case information and level of expertise. Memory studies replaced the earlier focus on general strategies as determinants of performance on medical diagnostic tasks (Bordage & Lemieux, 1991). Other studies have focused on a variety of knowledge structures, some of which have been derived from other research in Cognitive Psychology or created specifically for medical diagnostic problem solving. Also included are different "types" of knowledge.

Memory for Clinical Cases

Several studies have found no reliable differences between novices and experts on recall. For example, Muzzin, Norman, Jacoby, Freightner, Tugwell, & Guyatt (1982, 1983) found no overall difference in memory recall between experts and novices, which might be attributed to the experimental tasks used. However, they did find that experts recalled a smaller number of larger chunks than did novices.

In contrast, other studies have found marked differences in the information recalled at different levels of expertise. Two similar studies, found that on recall tasks, the information recalled by novices includes more verbatim case information and fewer inferences, while experts recalled more inferences and less verbatim (Coughlin & Patel, 1986; Patel, Groen, & Frederiksen, 1986). Patel et al. also found that students recalled and inferred many more irrelevant propositions than physicians did on an atypical case. The physicians' comprehension process was apparently different from students'; experts were filtering out irrelevant propositions. Finally, they conclude that medical problem solving is very complex, involving some interaction between knowledge, experience, and more "subtle" processes that develop out of experience and lead to diagnostic accuracy. Coughlin et al. also found that experts were able to selectively encode relevant and critical case information and gave higher-quality diagnoses.

Two studies by Norman, Brooks, Smith, and Henry (1987) investigated the recall of laboratory data. In the first, novice research assistants, 2nd year medical students, and experts worked on four case variations: 1) two levels of numerical data (actual patient data & random), and 2) two levels of organization (typical & scrambled). The amount of data recalled increased with expertise for the normal condition (patient data, typical format), but not for the random-scrambled condition. The experts apparently used a "highly configured search strategy" (p. 71) to look for related pieces of laboratory data. In non-normal conditions reading times increased for students and experts. In the second study, standard data presentation was used. Participants completed one of three tasks. The *formulation* task was to read the protocol, describe the problem, and link the laboratory results to the description. Students and experts recalled a small amount of data in this condition. The *incidental learning* task included the formulation task plus recall of specific laboratory values. The *intentional learning* task was to read, memorize, and recall data. Students recalled twice as much data in the intentional learning condition as in the incidental learning condition, while experts did the reverse. The researchers conclude that experts were using a qualitatively different strategy; the information they used was not simply memorized but worked into a problem formulation. They suggest that recall ability is simplistic for explaining how clinical reasoning evolves. It does not take into account information being stored in different ways at different expertise levels.

Recall was also studied in 2nd-year, 4th-year, and 6th-year medical students and internists (Van de Wiel, Boshuizen, & Schmidt, 1994a). Participants were randomly assigned to one of three time constraint conditions and asked to provide a diagnosis and a pathophysiological explanation for four cases. Diagnostic accuracy increased linearly with expertise. For pathophysiological explanations, they found main effects for level of expertise and processing time. The number of concepts generated increased from the shortest to the longest case exposure time. The total number of concepts generated was also higher for 6th-year students than any other group in all three time conditions, which they attributed to more elaborate processing. Task instructions, laboratory data and their impact on recall was investigated in another study (Van de Wiel, Boshuizen, & Schmidt, 1994b). The participants were 2nd year and 4th year medical students, and internists who were asked to study a case description, provide a diagnosis, and write down everything they could recall about the case. Half were told ahead of time about the written recall task. Accuracy was not affected by the experimental manipulations: it increased with level of expertise. An inverted-U relationship was found for recall and expertise level.

Comprehension

Over several years, Patel and colleagues have used Kintsch's (1988, 1992) *Construction-Integration Model* of text comprehension to examine medical problem solving (Arocha & Patel, 1995; Groen & Patel, 1988). This theory specifies two processes as components of comprehension, which Patel and colleagues have used as a basic for their hybrid symbolic-connectionist approach (Patel, Kaufman, & Arocha, 1995). *Construction* refers to the activation of concepts in text and subsequent elaboration, inferencing, and the assignment of strengths to concept-proposition pairs. This is followed by *integration*, or the spreading of activation to related concepts. The comprehension-integration cycle occurs as each unit (e.g., sentence) of text is processed. Developmental changes in how comprehension takes place have been found along with differences between novices and experts (Arocha & Patel, 1995a; Groen & Patel, 1988; Schmidt & Patel, 1987).

The relationship between comprehension and knowledge integration by individuals with four different levels of medical training was examined by Patel and Medley-Mark (1985). Participants were asked to read several texts of three different types: general knowledge (about cancer), inferential, and integrative. Comprehension was assessed based on the propositions recalled and inferred from text, using propositional analysis Frederiksen's (1975) framework. They found evidence of a developmental progression in ability to use rules. Specifically, their results indicated that the rules are learned in the following order: selection, generalization, and integration. They found that comprehension is a necessary but insufficient condition for integration.

A modified methodology for studying comprehension, the *explanation protocol* task, techniques was developed by Patel and Groen (1986) and used in several subsequent studies. Briefly, the new approach involved: 1) presenting participants with text-based case descriptions for a specified length of time and then removing them and subsequently asking participants to 2) recall as much of the case as possible, 3) to explain the pathophysiology, and finally 4) to provide a diagnosis. The researchers mapped propositional representations in the text and recalled information onto the pathophysiological representation, and found that all experts who made accurate diagnoses used forward reasoning, while those who were inaccurate used at least some backward reasoning. They also tentatively identified a network of causal rules that some of the physicians used in conjunction with forward reasoning to accurately solve problems. These causal rules were apparently developed from the participants' prior knowledge, and were activated by relevant propositions in the text presented to them.

Schemas

Adopted from other cognitive studies, one construct with explanatory value in terms of differences in diagnostic accuracy is the completeness or specificity of an individual's disease model or *schema.* Experts in studies by Feltovich, Johnson, Moller and Swanson, 1984 and Johnson, et al., (1981) had more complete schemas and stronger connections between them, and they accessed relevant knowledge more easily than novices. In another major study, expert schemas were found to be qualitatively different from those of novices and more efficiently invoked, both of which resulted in more effective performance than that of novices (Lesgold et al., 1988).

Perceptual Structures

Several studies point to the importance of visual data representations. Analogical representations (Mandler, 1998; McNamara, 1994; Rumelhart & Norman, 1988) are particularly important in some medical sub-domains including radiology and dermatology (Norman, Brooks, Coblentz, & Babcook, 1992). Some researchers have proposed that perceptual image data may be stored in declarative form and converted during use, but it is increasingly regarded as having a qualitatively different format (Kosslyn, 1996;

Kosslyn & Koenig, 1992). In one radiology study, Lesgold et al. (1988; Lesgold, 1990) explored the relationship between diagnostic outcomes and the characteristics of knowledge applied in the context of diagnosis. The findings included large differences in which structures subjects attended to and which concepts they applied. Experts created thorough representations of anatomy and analyzed overlapping objects on a film (x-ray). Other studies have also examined thinking in visual medical domains (Azevedo, 1997; Crowley, Medeveda, & Jukic, 2003; Norman, Muzzin, Somers, & Rosenthal, 1990).

Other Knowledge Structures

Expertise research has investigated various other knowledge structures. Drawing on research on categorization, Bordage and Zacks (1984) investigated the *prototype* or clear example of a given disease category. A prototype captures the core meaning or definition of the category it represents. These prototypes may act as a sort of indexing scheme that helps the physician in recalling other category members. Four experiments were conducted involving medical students and physicians who worked with various disease categories and exemplars with different levels of "prototypicality". Physicians had richer, more tightly networked knowledge based on a greater number of connections between the disorders within categories that they displayed. They were also faster in accessing their knowledge, which was attributed to greater associative strength between different disorders in memory.

In a study of the memory structures of medical personnel, Grant and Marsden (1987) had 1st and 3rd year medical students, senior house officers, registrars, and consultants complete four cases (two easy and two difficult). The experimental task was to: 1) provide up to five "ideas" or explanations for what might be wrong with the patient, and 2) identify what information or *forceful feature* led to each idea. The number of ideas and forceful features did not differ significantly across groups (however, the researchers limited the number of ideas to five per case), but the actual content of ideas differed significantly across groups in 3 out of 4 cases as did the forceful features in all cases. The researchers conclude that the content of thought differentiates levels of expertise, and that expertise develops based on "personal relevance and utility in relation to individualized memory structures" (p.97).

According to the *structural semantic perspective* (Bordage & Lemieux, 1991), data such as signs and symptoms can be viewed in terms of two levels. The elementary level refers to the contents or substance, while the deeper level refers to their form and abstract associations. Knowledge is represented in semantic networks, such that the meaning of a sign or symptom is understood in terms of its semantic relation to others in a set. The researchers analyzed think-aloud protocols from strong and weak students, and physicians. Successful diagnosticians (experts and students) were found to use more elaborated abstract relationships to solve problems, indicating they had deeper problem representations. Other research suggests that the structure and comprehensiveness of the *knowledge base* is critically important to medical problem solving (e.g.; Patel et al., 1986). Boshuizen (1994) found that level of experience is correlated with changes in the structure of the knowledge base and with the quality of diagnoses.

A *hierarchical framework for the structure of medical knowledge* (Evans & Gadd (1989) has been used in several studies. It identifies levels of medical knowledge that ranges in specificity from general to the "full understanding" of a particular problem. The *empirium* is the most general, defined as the "whole world" of possible observations about a patient. *Observations* are perceptual categories that provide the first level of an observable phenomena but may not be clinically useful. Clusters of observations or *findings* have explicit diagnostic potential. Clusters of findings called *facets* reflect aspects of a disease state. A *diagnosis* accounts for facets, findings, and observations and supports explanation and prediction. Finally, the *global complex* is the complete understanding of a case including short-term treatment and long-term interventions. Braccio (1988) adopted this hierarchical framework. In this study, 1^{st}-year students worked with observations (isolated pieces of data), 2^{nd}-year students had more available knowledge and used more sophisticated strategies, and 4^{th}-year students were better able to partition the problem space and build on facet-level representations. However, senior students lacked the knowledge to carry out data-driven reasoning, and used facet-level hypotheses to organize clinical information.

Finally, the *Small Worlds* hypothesis, which specifies a knowledge structure closely related to a schema, has been proposed by Kushniruk, Patel, and Marley (1998). According to this view, expert diagnostic knowledge is organized according to small worlds or sets of logically related diseases. When a patient problem is encountered, the expert compares the various diseases in a particular set or small world that corresponds to the case at hand. The researchers developed this hypothesis based on their work in medical artificial intelligence, specifically the expert system PATHFINDER. The researchers used this approach to re-examine data from previous medical expertise studies (e.g., Joseph & Patel, 1990). Consistent with this hypothesis, the HDK participants were very quick to generate and consider small sets of closely related hypotheses, and to focus on the specific features that would help them distinguish between competing hypotheses. Meanwhile, the LDK participants were unable to form small sets of possible hypotheses or to select hypotheses that could be verified based on available information. The researchers propose that a "semi-automatic, recognition-primed processing of familiar situations" (p. 268) is involved in expert medical problem solving.

Biomedical Versus Clinical Knowledge.

Several studies (Patel et al., 1986; Patel & Kaufmann, 1995; Patel et al., 1991) have investigated the role of *biomedical knowledge* in diagnostic problem solving (knowledge of basic biology and medical science). It is contrasted with *clinical knowledge* or knowledge of the characteristics of particular disease types. Study results indicate that experienced physicians use clinical knowledge and do not directly apply biomedical knowledge when diagnosing cases (Patel, Evans, & Groen, 1989). There is evidence to suggest that as expertise increases the biomedical knowledge applied becomes more tailored to the particular case (Lesgold et al.,1988). Research on biomedical and clinical knowledge is further discussed in the following section.

Illness Scripts: Knowledge Encapsulation

Based on the notion of a *script* as defined by Schank and Abelson (1977), an *illness script* (Feltovich & Barrows, 1984) is a knowledge structure that consists of: 1) *enabling conditions* that affect the probability of a person having a particular disease or condition, 2) a *fault* or major physical malfunction (biomedical knowledge), and 3) *consequences* or signs and symptoms (Custers, Boshuizen, & Schmidt, 1996). These scripts become more elaborated and expert-like through the key process of *knowledge encapsulation (KE).* Novices develop elaborated causal networks to explain the causes and consequences of disease in terms of underlying pathophysiology. With experience, their *biomedical knowledge* is compiled and stored as illness scripts. Physicians have illness scripts for every disease they are familiar with, and experience determines how "filled-in" an illness script is. Problem solving with routine cases is then a process of searching, selecting, and verifying scripts. Knowledge elaboration is involved and developmental-type shifts in the knowledge base lead to new conceptualizations. Schmidt and colleagues (Boshuizen, 1996; Schmidt & Boshuizen, 1991, 1993; van de Wiel, Schaper, Boshuizen, & Schmidt, 1995) have conducted a relatively large number of studies using varied methodologies. Their findings and conclusions concerning KE include the following.

- KE may play a role in the series of sudden developmental-type shifts in expertise development.
- KE is triggered by exposure to real patients and cases.
- KE entails a decrease in the use of biomedical knowledge with developing expertise (e.g., Custers, 1995; Custers, Boshuizen, & Schmidt, 1998; Hobus, Boshuizen, & Schmidt, 1989). Findings from think-aloud studies also indicate that little biomedical knowledge is used by experts, while students use it extensively (e.g., Boshuizen, Schmidt, & Coughlin, 1987). Boshuizen and Schmidt (1992) found that students produced more biomedical propositions and their knowledge was not as closely linked to their explanations as that of experts. The researchers concluded that experts use biomedical knowledge tacitly, because it is encapsulated in clinical knowledge. Hobus, Schmidt, and Boshuizen (1989) asked experienced family physicians and 6^{th} year students to describe the prototypical patient for a number of disease types. The less experienced participants included fewer enabling conditions in their descriptions. Both groups showed relatively low numbers of biomedical statements, possibly due in part to the relatively advanced stage of the students.

Rikers, Schmidt, & Boshuizen (2002) summarize their findings in relation to previous KE studies. Briefly, recall data is important according to the KE model as far as the form of recalled data. They note that studies by Patel and colleagues (e.g., Joseph & Patel, 1990) failed to find any differences between levels of expertise on recall tasks, and attribute the differences to variations in analytic techniques rather than the use of *generic* versus *specific* expertise as suggested by Patel and colleagues. They suggest that KE is likely not the only factor influencing the non-experts' use of biomedical knowledge. In some studies (e.g., Lesgold et al., 1988) experts explicitly used biomedical knowledge (although this finding may be restricted to radiology). Since the overall number of encapsulated concepts recalled was relatively low (17% for experts and 10% for non-experts), they conclude that other factors are probably involved.

A number of additional findings, briefly reviewed here, have been obtained from studies designed to examine illness script development and use in medical diagnosis.

- Less-experienced participants often try to account for nearly all cues, while expert explanations were highly encapsulated (Schmidt, Boshuizen, and Hobus, 1988).
- Experts produce significantly more accurate initial hypotheses than students, and the recall more segments than students (especially relevant information) (Hobus, Schmidt, Boshuizen, & Patel, 1987). However, in the Rikers et al. (2000) study, the number of recalled propositions and the number of concepts used to explain the signs and symptoms both fell into the inverted U-shape. Experts and 2nd-year students performed at about the same level, while 6th-year students recalled/used significantly more.
- Experts give higher quality, more elaborate, more complete, and more fluent case explanations (Van de Wiel, et al., 1995). In addition, Rikers et al. (2000) found a linear increase with expertise for both the number of concepts used that matched concepts in a canonical model, and the number of links made between concepts.
- Enabling conditions apparently become increasingly integral components of illness scripts; more are mentioned with increasing expertise (Custers, 1995).
- Typical information is processed faster than atypical information and experts are very sensitive to combinations of prototypical enabling conditions and consequences. Custers, et al. (1996) found that experts were significantly faster at reading case descriptions than the students, which the researchers attribute to experts having well-formed illness scripts that allow for rapid processing.
- Experts are significantly faster and more accurate than students in completing case explanation tasks, both measures increasing linearly with expertise (Rikers, et al., 2000).
- All participants process cases faster and more accurately in their domain of expertise, but apparently, specialists do not process cases in a "qualitatively" different manner within and outside their domains of expertise (Rikers, Schmidt, Boshuizen, & Linssen, 2002). In addition to being more accurate, experts working in their domain of expertise recall more encapsulated concepts and are better at focusing on the important aspects of the cases. Experts outside their domain of expertise recall more propositions in a literal format and mention more non-essential aspects of cases (Rikers, Schmidt, & Boshuizen, 2002).

Characterizing the Research

Summary of Findings

Based on this literature review, some characteristics of individuals with different levels of expertise can be tentatively stated (see *Methodological Concerns*). Each group, "novices", "intermediates", and "experts", in these studies and in the following summary actually encompasses different levels of training.

- Typically, there is a linear increase with expertise in diagnostic accuracy, and a linear decrease in case solution and reading time. Experts can take longer to solve or read cases, which has been attributed to more elaborate processing and explanation. This finding breaks down for atypical cases and outside the domain of expertise.
- The intermediate effect has sometimes been explained by novices' use of obvious or prototypical knowledge, while intermediates have difficulty sorting out what is relevant, dealing with time constraints, and excluding incorrect hypotheses. Non-experts often try to account for nearly all cues, while experts do not.
- Experts can efficiently identify correct diagnoses, rule out others, and confirm diagnoses. Non-experts and experts outside of their domain continue to generate hypotheses, and cannot effectively rule them out.
- Experts efficiently eliciting information in patient interviews, and they use it to develop integrated knowledge structures. Intermediates vary, and may not use evidence to rule out hypotheses. Novice data-gathering is inefficient, they do not give accurate diagnoses, and their knowledge may be superficial or non-existent.
- At all levels, diagnostic hypotheses are generated in response to case history information including the "chief complaint". To a large extent, experts generate the correct diagnosis in response to this data, while novices and intermediates do so to a lesser degree.
- Limited data suggest that experts may use more planning in diagnostic problem solving than non-experts.
- Experts recall less data verbatim, and make more inferences and novices do the opposite. Experts also recall more encapsulated concepts than non-experts. The amount of data recalled by expertise level shows conflicting findings.
- Experts' knowledge structures (scripts, schemas, prototypes) are more complete and accessible, and interconnected than those of novices.
- As novices become more experienced, they progress from working with isolated pieces of data to using clusters of findings and more sophisticated strategies.
- Experts quickly generate and examine small sets of closely-related hypotheses, while residents are unable to do so.
- Biomedical knowledge is used by novices extensively. Experts use clinical knowledge, which incorporates encapsulated biomedical knowledge.

- Experts also give faster, higher quality, more elaborate, more complete, and more fluent case explanations than non-experts.

These findings together point to a developmental course with specific characteristics. Early in the developmental course, knowledge is limited, it takes more time to complete tasks, all cues are potentially important, hypotheses are difficult to rule out, data gathering is inefficient, knowledge structures are weak, isolated pieces of data rather than clusters are examined, and biomedical knowledge is used extensively. As development proceeds, more knowledge is acquired, but it is still difficult to use effectively which can result in poor performance compared to both novices and experts. Mixed strategies are often used. Eventually, development results in rich expert knowledge structures and the ability to perform efficiently and effectively using strong problem solving methods. For non-experts and experts outside of their domain, only certain sub-components of expertise are present, and diagnostic problem solving is less accurate and less effective in various ways.

So far, work in this area has not established a single approach to account for medical expertise. Rather, these various approaches seem to each have some applicability and they have often established common findings, presumably because they are general enough to map onto this complex cognitive processing.

Methodological Concerns

Medical expertise research, taken together, has some general characteristics that are important. This review has demonstrated that, beyond the shared general cognitive orientation and the goal of identifying differences in terms of the cognitive processing of individuals with different levels of experience, in certain significant respects there is in fact little consistency between the studies. Researchers have included cognitive scientists interested in different topics (e.g., cognitive modeling, reasoning) as well as physicians from various specialty areas with various interests (e.g., medical teaching, PBL). These researchers have directed their efforts towards different research goals, frequently the investigation of specific cognitive factors. As mentioned, many of the studies have employed one of two standard methodologies, often modified for specific purposes.

Further, diagnostic problem solving is obviously complex, but it also seems to vary across tasks in the domain. Research efforts have not found a single structure or process to *explain* medial problem solving, but they have shed light on different components of processing. However, the findings very often lack generalizability, due to methodological variations and the particular sub-domains of medicine they are conducted in. Medical expertise research therefore does not provide a coherent body of work; the results of particular studies may or may not prove true if replicated, for example, in different sub-domains of medicine. Indeed, conflicting findings have been reported. The review provides some understanding of these varied research efforts and conclusions drawn based on them. The various results can be seen as compatible with a wide range of proposed structures and processes (Norman, 2000).

A number of methodological concerns associated with medical expertise studies create potential problems in terms of generalizing the results. Indeed, the literature has produced conflicting results on several measures, and some researchers have cited methodological inconsistencies as potential explanations. The following list identifies some of the concerns.

- *Defining problem solving.* One methodological concern is inconsistency in what diagnostic "problem solving" refers to (Boshuizen & Claessen, 1982). Many studies have focused specifically on the task of diagnosis. Others have examined related activities that do not share the same features.
- *Defining levels of expertise.* Another problem area is the definition of various levels of expertise. Across studies, the characteristics of novices, intermediates, and experts have varied, although some researchers have attempted to provide general definitions (e.g., Elstein et al., 1990; Patel, Arocha, Kaufman, 1999b). For example, 3rd year medical students have been labeled *novices* in some studies, and *intermediates* in others. Other levels of expertise have included *experts working outside of their domain*, *basic experts* and *super experts* (Raufauste et al., 1998), *non-experts, sub-experts, lay people, beginners* (Patel & Groen, 1991b), *good students, poor students* (Bordage & Lemieux, 1986), and *early, intermediate*, and *advanced novices* (Arocha & Patel, 1995).
- *Varied, non-authentic experimental tasks and "laboratory-type" settings.* These studies have employed various experimental tasks, which in turn elicit thinking that resembles medical diagnostic thinking to varying degrees. Typically, the tasks have been not been "authentic" as far as the thinking involved. In actual diagnostic contexts, a physician actively elicits information (a case history, a physical examination, and diagnostic tests). Experimental tasks have almost always involved participants working with information already summarized for them. Studies have also largely been conducted in laboratory settings rather than more naturalistic hospital settings. Compared to many studies, the current study used a somewhat more realistic diagnostic task.

- *Concerns with cases.* Another concern is the use of simplistic cases that differ from the complex patients often encountered in real medical practice. Several studies have manipulated cases to make the data inconsistent, with the goal of examining the impact on problem solving at different levels of expertise.
- *Sample size.* Many studies have used small numbers of participants, even a single individual per level of expertise. In the current study, two reasonably large groups of participants at each level were included (10 per group).
- *Various medical sub-domains.* Further complicating the matter, these studies have also been conducted in different specialty areas that are not necessarily uniform in terms of the cognitive processing involved. The distinction between visual and non-visual medical domains (e.g., radiology versus internal medicine) has been established, In visual domains, problem solving may rely fundamentally on pattern recognition, while other forms of data would be elicited and processed in non-visual domains. Beyond this general distinction, research has not emphasized the features of particular sub-domains.'

In addition to methodological concerns, some researchers take the fairly general results as evidence of only limited theoretical advance in the study of expertise. Schmidt and Boshuizen (1993) cite three reasons: largely descriptive research, quasi-experimental designs as the preferred method of investigation, and an apparent lack of stable empirical phenonema in the field.

Summary

This portion of the literature review has provided an overview of expertise research, with a reasonable detailed look at medical expertise research over the past few decades. Several important findings have been emphasized. Together the results from studies of a fairly wide range of factors (general measures, cognitive processes, and cognitive structures) have shown that, from many different approaches, this research has established many common findings. The remainder of this review focuses on learning environments for medical education, including both implications drawn by medical expertise researchers, and, in keeping with the goals of this study, computer-based learning environments for medical education.

Learning Environments for Medical Education

Implications from Medical Expertise Research

As noted by Alexander (2003), findings from expertise research in general have not translated directly into educational practice, partly because the studies have not been conducted in school settings with students in mind. Education has just not been the focus of this research. Accordingly, medical expertise researchers have typically not explicitly used their results to make recommendations for training. A suitable theory of learning and instruction needs to account for skilled performance, acquisition, and instructional intervention (Patel, Glaser, & Arocha, 2000). When recommendations have been made, they have typically been very *general,* as follows:

- Patel (1990) suggests that instructors should have a good understanding of *students' prior knowledge*. Further, *assessment should look at knowledge structures* (concepts and the connections between them).
- Greater instructional emphasis should be placed on *explanation* in learning (Groen & Patel, 1991; Patel & Groen, 1991a).
- Medical education should emphasize 1) *problem-oriented teaching and learning* (rather than science-centered approaches), 2) *awareness of types of reasoning* (Magnani, 1992), and 3) *the use of computer systems that allow for explicit distinction between knowledge and reasoning processes.*
- Elstein and Schwarz (2002) recommend PBL as 1) a means for introducing hypothesis generation and testing into medical curricula, and 2) preferable to learning general strategies. Also, *biomedical and clinical knowledge should be examined in relation to different instructional models*, e.g. PBL (Patel & Kaufman, 1995).

While potentially useful, the generality of these recommendations speaks to a need for studies that can advance development of instructional principles and detailed recommendations for medical education. As Patel et al. concede "we still need to understand more about the conditions of learning that lead to more optimal levels of performance" (p. 259). Studies are needed that focus on particular learning environments and use realistic diagnostic tasks. In keeping with the goals of this research, the remainder of this review focuses on technology-based learning for medical education.

Computers and Learning in Medical Education

Many researchers in different fields suggest that computer-based instruction is a promising direction for medical education (Lajoie 2003; Patel, Arocha, & Kaufmann, 2001; Patel & Kaufmann, 1998; Patel, Kaufmann, & Arocha, 2002). While medical expertise research does not translate directly into detailed implications for computer-based instruction, other research in medical education, medical informatics, and educational psychology is relevant for the development of CBLEs for medicine. This brief review provides a context for applying the results of the study to the development of BioWorld.

Learning Technologies

Over the last several decades, several types of learning technologies have been developed that link theories of learning to the design of instruction (Lajoie, 2000). Included have been early programmed instruction, computer-assisted instruction (CAI), computer-based training (CBT), intelligent tutoring systems (ITS), expert systems, simulations, hybrid systems, and CBLEs. As was the case over thirty years ago (see Clancey, 1986), computer-based instruction is still touted as a promising direction for the implementation of learning environments (Chipman, 2003; Lajoie, 2000). In recent years, advances in both learning theory and technology have continued to increase the design options for CBLEs (Bruer, 2003; CTGV, 2003; du Boulay & Luckin, 2001).

The Need for Principled Design: Student-Centered CBLEs

For decades researchers have suggested that current learning and instructional principles should be incorporated into the development of computer-based instruction (Chan & Baskin, 1988; CTGV, 1996; Hativa & Lesgold, 1991; Mayer, 2003; Winn, 2002). The term *principled design* refers to learning activities that are designed based primarily on some form of learning and/or instructional theory (Lajoie, 2000; Lajoie & Derry, 1993). These designs are created with the goals of 1) engaging learners in particular forms of cognitive processing and thereby 2) encouraging certain learning outcomes that are desirable based on the theory. This approach to design differs from other approaches commonly used, say, in non-academic development contexts. For example, many systems are designed around a specific technical capability, perhaps with learning as a secondary consideration.

A general principle derived from Constructivist and Cognitive theory says that active learning or learning by doing is important for constructing one's own knowledge and making learning meaningful (Lesgold & Nahemow, 2001). While designs vary, the general goal *of student-centered learning environments* is to promote learning through realistic activity in a domain, with the student directing the activity (Land & Hannafin, 2000).

Learning Factors to Support in a CBLE

Recently, student-centered learning CBLEs have been designed to support a range of cognitive and affective factors related to learning. Some examples include systems for supporting *metacognition* (Aleven & Koedinger, 2002; Puntembekar & du Boulay, 1997), *reflection* (Katz, O'Donnell, & Kay, 2000; Kolodner & Guzdial, 2000), *motivation* (Del Soldato, 1994), *self-regulation/ monitoring/ evaluation* (Atkinson, Renkl, & Merrill, 2003; Hadwin, 1996; Ley & Young, 2001; Mayer, Dow, & Mayer, 2003), *transfer* (Bassok, 2003), *decision making,* and *collaboration* (Kreijns, Kirschner, & Jochems, 2003).

Instructional Models, Strategies, and Assessment

Instructional models. A range of CBLEs have been developed, generally in academic research contexts, based on cognitive learning principles and models of instruction and the student-centered approach. The environments of interest are designed to engage learners in working on "authentic" activities within a particular domain. Recent learning and instructional "theories" such as the various versions of *constructivism* and *situated learning* (Lave & Wenger, 1991), along with corresponding instructional models have been used as the basis for models the design of a range of CBLEs (Jonassen & Land, 2000). Included are discovery learning, PBL, cognitive apprenticeship, microworlds, anchored instruction, project-based environments, inquiry-based environments, simulations, and collaborative learning environments. Many of these learning environments are specifically designed to support problem solving. (e.g., Goldman, Zech, Biswas, Noser, Bateman, Bransford, Crews, Moore, Nathan, & Owens, 1999; Lillehaug & Lajoie, 1998).

Cognitive Apprenticeship (Collins, Brown, & Newman, 1989) provides a model for *scaffolding* learning by providing support as needed to assist learners as they perform *authentic* tasks. It has been investigated in schools, medical education classrooms (Lajoie, Faremo, & Wiseman, 2001), training settings (Lajoie & Lesgold, 1989) and CBLE contexts (Cho & Jonassen, 2002; Lajoie et al., 2001; Saye & Brush, 2002).

A number of *Discovery Learning* environments have also been developed. De Jong, and Van Joolingen (1996) address issues relating to discovery learning environments that involve computer simulations of conceptual models, including characteristics of simulations and learners. They also provide a lengthy discussion of issues relating to instructional support for learning.

Instructional strategies. Other research that is important to the current study includes work on tutoring and instructional strategies and models that can be implemented in CBLEs for medical training (e.g, Lajoie, et al., 2001). For example, Du Boulay and Luckin (2001) raise a number of issues relating to computer-based tutors and human teaching strategies, including the difficult topic of what instructional strategies should be implemented in computer-based learning environments, focusing their discussion on how to deal with errors and how to motivate students.

Assessment. Various researchers have addressed the need for new forms of assessment for learning from problem solving (e.g., Baker & O'Neil, 2002) and other "authentic" activities. New forms of assessment in medicine are also in demand (Patel, 1990), and new assessment models have been proposed (e.g., Charlin, Brailovsky, Roy, Goulet, & van der Vleuten, 2000).

In summary, computer-based instruction remains a very active area of educational research. Recent work is wide-ranging in the scope of environments and types of thinking they are designed to support. Over a decade ago, Elstein and Rabinowitz (1993) suggested that medical cognition research has begun to affect medical training. Some researchers have now designed CBLEs for medical training, and the remainder of this literature review will briefly review these systems. This in not an exhaustive review; rather, it provides an overview of important topics and research.

CBLEs to Support Diagnostic Problem Solving in Medicine

The various types of CBLEs described in the previous section have also been developed for different areas of medical training, from teaching diagnostic skills to decision support and patient management (Anderson, 1992; Clancey, 1988; Elliot, Williams, & Wolff, 1996; Lajoie, 2003; Lillehaug & Lajoie, 1998; Shortliffe, 1991).

Early Efforts

Focusing on systems to support the acquisition of diagnostic skills, a few early examples demonstrate the range of approaches for computer-based systems to support the acquisition of diagnostic skill in medicine. One system developed by Warner, Woolley, and Kane (1974), was designed to teach medical students to solve problems in relation to history taking. The goal of the system was to help students to organize their thinking while solving patient problems, by applying formalized logic (Bayesian probabilities) to patient problems. The system required students to make several tentative diagnoses during a case, which would then be compared to the results generated by the computer. During the same decade, Wortman (1972) used IPT to develop a computer program that could actually "conduct" or model medical diagnosis.

During the 1980's research focused on rule-based expert systems and later on intelligent tutoring systems or ITSs (Lesgold, Eggan, Katz, & Rao, 1992). A well-known example, the knowledge base from the rule-based expert system MYCIN (Buchanan & Shortliffe, 1984; Clancey, 1984, 1988) was used to develop GUIDON, an intelligent tutoring system designed for medical education. It used the knowledge base from MYCIN. After some time, the two systems were modified and re-named as NEO-MYCIN and GUIDON2. Several levels of clinical problem solving were identified in keeping with current expertise research.

More than a decade ago, the *Medical Center* (Anderson, 1992) was developed for obstetrics and gynecology. It utilizes protocols of expert clinician-patient interviews as "the basis to support learner access to expert clinical reasoning and associated webs of basic science knowledge" (p.372). The system includes real-time videotaped expert-patient interviews based on real patient records. Several experts together developed a reference matrix of clinical findings and hypotheses that represent the full search space. This matrix is used as the basis for supporting learner exploration of the diagnostic process and the associated knowledge structures.

Computer-Supported Collaborative Learning (CSCL)

CSCL is noteworthy as a model of instruction for computer-based instruction (Koschmann, 1996; Koschmann, Kelson, Feltovich, & Barrows, 1996). It is a collaborative learning model based on sociocultural theories of learning. Koschmann (2001) reviews several paradigms of instructional technology, and gives support for CSCL because it "has the advantage of studying learning in settings in which learning is observably and accountably embedded in collaborative activity" (p. 19). According to Koschmann, CSCL is consistent with various learning theories, from social constructivism, to activity theory, and situated learning, but the CSCL research community has yet reach to reach consensus on what learning is and how to study it.

Student-Centered Learning Environments for Medical Training

Many CBLEs have recently appeared in the medical education and medical AI literature, including simulations, agent-based simulations, hybrid systems, collaborative problem-based practice environments sometimes delivered via the Internet and World-Wide Web (WWW) (e.g., Bowdish, Chauvin, Kreisman, & Britt, 2003; Kim, Kolko, & Greer, 2002; Murphy, Friedman, Elstein, Wolf, Miller, & Miller, 1996). The most relevant systems for the purposes of study are designed to support diagnostic problem solving in authentic, often simulation-type environments..

Learning Through Simulation CBLEs

Two different CBLEs have been developed for training in surgical intensive care unit (SICU) nursing. The SICUN is a simulation-based ITS (Lajoie & Azevedo, 1998; Lajoie, Azevedo, & Fleiszer, 1998), designed based on situated learning principles and the cognitive apprenticeship framework (Collins et al., 1989). It provides users with several cognitive tools to support complex decision making. The researchers examined the rapid clinical decision making of expert SICU nurses and found great variability in how they made clinical decisions, although they achieved the same final results. Differences were also found in hypothesis generation, planning, heuristic use, and overall solution paths. Alexe and Gecsci (1996) describe *SICULE,* another CBLE that can adapt its learning environment to allow for different learning modes, including exploration, demonstration, and task resolution.

Intelligent Agent-Based Simulations

Two notable examples of agent-based simulations will be described. First, Ganeshan, Johnson, Shaw, and Wood (2000) describe *Adele*, which was designed for use in undergraduate case-based clinical instruction. The system allows students to examine simulated patients, collect data about medical history, conduct a physical examination, request diagnostic tests, and make diagnoses. As students work with cases, the intelligent agent called Adele monitors their progress and provides feedback. A Bayesian network is used to represent causal relationships between hypotheses and findings, which in turn are used as a basis for its tutoring dialogue. Student knowledge is probed based on network structure. Adele can then provide hints for how to proceed, engage students in learning and reasoning about underlying pathophysiology, provide feedback, and initiate different forms of dialogue in response to student actions that are correct or incorrect.

Another agent-based environment, *LAHYSTOTRAIN* (Los Arcos, Muller, Fuente, Orue, Arroyo, Leaznibarrutia, & Santander (2000) is a prototype designed for training two surgical techniques: laparoscopy and hysteroscopy. It supports the acquisition of content knowledge and practical skills; learners first acquire knowledge of the techniques, and then practice them using virtual instruments and a virtual patient. A number of behavioral agents (nurse, anaesthetist, and surgeon) assist the learner. A pedagogical agent, the Tutor, manages the learning environment, deciding which lessons and exercises to recommend based on the level of experience of the learner. The system also has a user model, which can provide low, medium, or high support.

Other Systems

A model-tracing Intelligent Tutoring System, *SlideTutor*, is designed to create training systems for diagnostic classification (Crowley, Medvedeva, & Jukic, 2003) Central to the design are rule-based production systems. Currently, the system contains domain knowledge for dermatology, a domain characterized as involving feature-based classification. The knowledge representation components of the system consist of a domain model, case data, and a domain task model that represents goals. Their Dynamic Solution Graph is designed to generate valid paths through a problem space, based on the contents of the knowledge representation components. Finally, an Instructional Layer provides feedback during problem solving.

The Problem List Generator (PLG) is designed to assist clinical pathology students in veterinary medicine to learn from diagnostic problem solving (Danielson, Bender, Mills, Vermeer, & Lockee, 2003). PLG learning activities are similar to BioWorld's: 1) review the equivalent of case history and physical examination information, 2) review laboratory tests and results, 3) mark any case information that is abnormal, 4) construct a hierarchical *problem list* of data items, 5) identify the "mechanisms" of disease based on biomedical knowledge, 6) make a diagnosis, and 7) compare results to an expert problem list. Danielson et al. studied the effectiveness of the PLG and found: higher final examination scores, positive participant and instructor opinions, and positive ratings for usability. Finally, they draw two general recommendations for design. The PLG was effective because 1) students identified all of the abnormal or pertinent case information before building their problem lists, and 2) a common format for presenting one's understanding of a problem was a "powerful feedback mechanism" (p.73).

Summary

This section has provided an overview of some of the research conducted in CBLEs in general, and for medical education, specifically training medical diagnosis. Researchers have taken various approaches to the design of learning activities for these systems. The importance of the principled design of students centered learning environments has been highlighted. Both the design of these systems and the results they have achieved with medical students are potentially informative for the ongoing development of BioWorld.

Chapter Summary

Medical expertise studies have investigated many aspects of diagnostic problem solving in recent decades. As summarized by Custers et al. (1998), the primary difference between experts and non-experts is the quality of their relevant knowledge, rather than other factors that have been examined (e.g., general problem solving ability or diagnostic strategy use). In addition, this superior quality of knowledge has an influence on performance throughout the diagnostic process. Finally, for a number of performance measures, experts and novices are not consistently differentiated. That is, performance does not always improve in a linear manner with increasing expertise. Several methodological concerns were also highlighted, some of which are addressed by the current study. A brief review of research on computer-based learning environments was also included to provide a context for the current study. The following chapter describes the methodology used to conduct this study.

CHAPTER 3: METHOD

Research Questions

The following research questions were addressed in this study.

1. How does the general performance (time on task, protocol length, diagnostic accuracy) of students and residents differ as they attempt to diagnose internal medicine cases based on case history, physical examination, and diagnostic testing activities? How does their performance compare with that of experts?
2. How does the problem solving of students and residents differ as they attempt to diagnose internal medicine cases based on case history, physical examination, and diagnostic testing activities? How does their problem solving compare with that of experts?
3. How does the post-problem reflection (categorizing data, prioritizing data, and comparing to an expert list) of students and residents differ? How do their réflective actions compare with those of experts?

Participants

The participants consisted of 10 third-year medical students and10 residents enrolled in the Faculty of Medicine at a Montreal university. All of them were on rotation at a university teaching hospital. The groups were chosen because they represent the range of medical knowledge and expertise for potential BioWorld users. At the time of data collection for students, they were approximately 3 months into their third year of study. Two Internal Medicine experts also participated. At the time of this study, undergraduate medical training at the university covered the following general topics:

- Year 1: Basis of Medicine - Basic Sciences
- Year 2: Basis of Medicine (6 months), Clinical Work (6 months)
- Year 3: Clerkship - Practice of Medicine
- Year 4: Electives (primarily applied to clinical medicine)

The residents had completed their four-year undergraduate training and were enrolled in residency training (in various medical specialty areas). They were primarily first-year residents, although a few more senior residents were also included.

Based on the Pre-Questionnaire (Appendix A) data, the average age of the third-year students was 24.6 years, and they reported having between 20 hours and 8 weeks of classroom instruction on the gastrointestinal system (the cases used in the study are all from Internal Medicine). The residents had an average age of 29 years and were specializing in various medical areas (e.g., anesthesiology, general surgery, radiology).

Four medical experts also participated in the study. Each of them had over 20 years experience practicing and teaching medicine. Expert 1, a surgeon, assisted in recruiting participants from one of the university's teaching hospitals. Expert 2, an Internal Medicine expert and an award-winning medical teacher, participated in several aspects of the study including the recruitment of subjects from a second teaching hospital, the initial re-design of the cases, the integration of the expert models, and data analysis. Finally, experts 3 and 4 both specialized in Internal Medicine and provided the protocols used to develop the expert models for the cases.

Materials

The Cases

Four existing internal medicine cases from BioWorld were selected for use in the study. One was used as a practice case. Each of the cases is a disease or condition affecting the gastrointestinal system. The researcher worked with Expert 2 in order to refine each of the cases, which involved eliminating minor inconsistencies and altering the content of the cases to better match the level of the medical students.

During data analysis, the expert was asked to rate the level of difficulty for each case. The instructions were to "rate the level of difficulty for medical students to determine the *correct diagnosis* for that case", with the correct diagnosis as specified in Table 1 below. A 5-point Likert scale was used for rating the cases (1= easy, 2 = moderately easy, 3 = intermediate, 4 = moderately difficult, and 5 = difficult).

For each case (except the practice case), Table 1 provides: 1) a disease description, the 2) "correct" diagnosis based on the available data, and 3) Expert 2's rating of the level of difficulty. The disease descriptions are based on relevant medical literature (Berkow & Fletcher, 1987; Spivak & Barnes, 1990), and they provide general information to support understanding of these specific cases. These descriptions do not provide comprehensive information about the disease types.

Table 1.

Case Descriptions, Diagnoses, and Level of Difficulty

Case	Disease Description, Case Information & Diagnosis	Difficulty
#1 Celiac	***Disease Description*:** A chronic intestinal malabsorption disorder caused by intolerance to gluten. It presents in a variety of ways, and some direct clues (diarrhea, abdominal discomfort, and distention) may be present. Iron deficiency is found in adults, and there can be 1) low albumin, Ca, K, Na, and 2) elevated alkaline phosphotase and prothrombin time. Diagnosed by symptoms, lab studies, and x-rays (Anti-Gliadin Antibodies is "the gold standard" test for detecting this disorder). ***Case Information*:** 27 year-old male, 2-year history of intermittent diarrhea, weight loss, and anxiety ***Diagnosis*:** *Malabsorption*, Celiac possible but, lacking confirming evidence (e.g., Antigliadin Antibodies).	(5 = difficult)
#2 Shigellosis	*Disease Description*: Shigellosis is an acute bacterial infection of the bowel, spread by Shigella organisms (4 types). It is spread directly by fecal-oral transmission, and indirectly through contaminated food or objects. In young children, there is sudden onset with nausea and vomiting, irritability, anorexia, and abdominal pain. Within 3 days, blood, pus and mucus appear in the stools, and the frequency of diarrhea increases. Bacillus is found in stools, hemoconcentration is common, and plasma CO2 is low (metabolic acidosis). Stool culture shows "sheets of leucocytes". Differential diagnosis should include Salmonella, Yersinia, Camphylobacter, amebiasis, and viral diarrheas. *Case Information*: 4-year old male, fever & fussiness, 4 days of bloody diarrhea, possible fecal-oral contact, previous antibiotics & aspirin *Diagnosis*: *Infection (bacterial)*, with specific possibilities: *Shigella, E-Coli, Salmonella,* etc.	(3 = medium)
#3 Hepatitis A	***Disease Description*:** The Hepatitis A virus (HAV) spreads mainly by fecal-oral contact. It may be spread by food and water, especially in under-developed countries. Hepatitis causes flu-like symptoms (nausea, vomiting, anorexia, and malaise), and later dark urine and jaundice. It affects the liver: SGOT and SGPT are elevated, and alkaline phosphotase is mildly elevated, WBC (white blood cell) count is low-normal, and a blood smear often shows atypical lymphocytes. A Hepatitis A titer detects the presence of this virus. ***Case Information*:** 25 year-old female with recent travel history, fever, abdominal pain, and jaundice. ***Diagnosis*:** *Hepatitis A*, typical symptoms, confirmation possible by positive Hep A titer.	(1 = easy)

*(1= easy, 2 = moderately easy, 3 = intermediate, 4 = moderately difficult, & 5 = difficult)

Case History and Testing Information Relevant to the Diagnoses

Table 2 presents a summary of the Case History, Patient Chart, and Testing information that is relevant for diagnosing each of the cases. It is important to note that in BioWorld all of the available Case History and Patient Chart information is presented directly to the participants. In contrast, they must make requests (e.g., order diagnostic test results) in order to gather Testing information. Therefore, the individual participants only had access to the Testing information that they specifically requested.

Table 2.

Case History and Testing Information Relevant to Diagnoses for the Cases

Case 1: Celiac (difficulty rating: Difficult)	
Case History	27 year old systems analyst, intermittent mild diarrhea for 2 years, seen several doctors, recent breakup and anxiety, lost weight over past 4 months, fatigued, 1-2 martinis a day, lately nocturnal diarrhea
Patient Chart	Vital signs within normal range
Testing	Negative stool culture, low test results for: folic acid, carotene, % SAT transforrin, ferritin, and serum iron
Case 2: Shigellosis (difficulty rating: Medium)	
Case History	4 month old, experiencing fever & fussiness, has been given aspirin, severe diarrhea and vomiting, bloody diarrhea, after 4 days still has a fever, suggestion of fecal-oral contact, previous antibiotics
Patient Chart	High temperature, other vital signs at the high end of normal range
Testing	Negative stool culture for Clostridium Difficile, elevated hemoglobin, hematocrit, & WBC, dehydrated on physical exam
Case 3: Hepatitis A (difficulty rating: Easy)	
Case History	Recent return from Mexico, nausea and vomiting, possible fever, slight abdominal pain, no appetite, fatigued, discolored urine (dark), yellowish skin discoloration around lips and eyes
Patient Chart	Slight temperature, elevated pulse
Testing	Increased indirect bilirubin, increased SGOT, SGPT, & alkaline phosphotase, increased WBC especially lymphocytes, positive Hepatitis A titer, negative Hepatitis B & C titers, negative mono screen

In addition, it should be noted that Table 2 contains only the Testing information that experts found important for diagnostic purposes (based on their protocols). This information consists of abnormal results (e.g., elevated temperature), as well as "pertinent negatives" or test results that are normal. Both abnormal or positive test results and these pertinent negatives can help to narrow a list of possible diseases. See Appendix B for the full Case Histories and Physical Exam information.

The BioWorld Learning Environment

The BioWorld learning environment is organized around two phases, each of which includes three learning activities. Learners first complete the *problem solving phase*, followed by the *post-problem reflection* phase. Each learning activity is described below. For additional information, see Lajoie, Lavigne, Guerrera, and Munsie (2001).

Task Analysis

A task analysis was conducted to identify the steps involved in completing BioWorld cases and the results are presented in Figure 1. The task analysis involved working through several cases to identify the BioWorld phases, the learning activities, and the generic sub-tasks required to complete each activity. Learners typically proceed

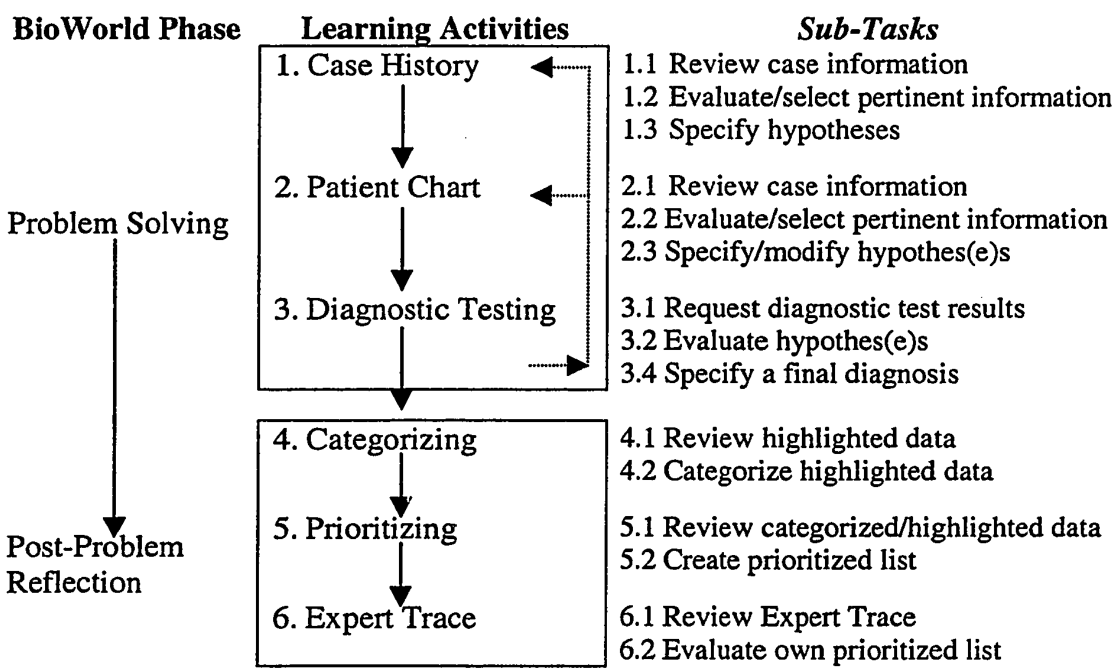


Figure 1.

Task Analysis: BioWorld Phases, Activities, and Sub-Tasks through the *Problem Solving* activities, followed by the *Post-Problem Reflection* activities, in a linear manner (Tasks 1-6, in order). They may or may not carry out all of the sub-tasks listed in the figure for each activity. During the Problem Solving phase, the order of activities may be varied to a limited extent. For example, a learner could revisit the Case History after viewing the Patient Chart, or access the Library. The user posts a final diagnosis to complete the Problem Solving phase and s/he is then directed to proceed with the Post-Problem Reflection phase. In this study, a few participants asked if they could return to problem solving activities (e.g., to request an additional test). They were permitted to do so, as they would be able to do in the BioWorld system.

Problem Solving Phase

1. *Case History.* During the first activity, the learner reads a short text containing a case history (symptoms, current complaints, etc.) and extracts relevant information. Evidence is collected by double-clicking on it to enter it in the *Evidence Palette.* The learner can enter diagnostic hypothesis from a pull-down menu for *Current Hypothesis* at any time, and adjust the *Belief Meter* to indicate how certain they are of their diagnosis.
2. *Patient Chart.* During this activity, the learner reviews patient information based on a simulated patient encounter upon admission to hospital. The chart contains: 1) a reason for admission to hospital, 2) patient age and gender, and 3) vital signs (temperature, blood pressure, respiratory rate, and heart rate). The learner reviews the information, and then has the option of adjusting their diagnostic hypothesis using the *Select Hypothesis* pull-down menu. Once a hypothesis has been selected, the user can then adjust the *Belief Meter.*
3. *Diagnostic testing.* During the last problem solving phase activity, the learner orders as many diagnostic tests, in any order, as s/he needs to solve the problem. The requested test results are presented, along with the expected or normal values. When the testing has been completed, the learner must enter a correct diagnosis to proceed to Post-Problem Reflection. If an incorrect diagnosis is entered, a feedback message will tell the learner that his or her diagnosis is incorrect. S/he must then attempt to revise the diagnosis.

Post-Problem Reflection Phase

4. *Categorizing Evidence.* The fourth BioWorld activity involves categorizing the information that was used to make a diagnosis into types (case history, diagnostic test results, etc.). The contents of the Evidence Palette are presented to the learner in a list, and s/he highlights the items one by one, and then selects the category in which each item belongs. Any information that is not selected is subsequently discarded by the system. When this activity has been completed, the learner clicks on the *Next* button.
5. *Prioritizing Evidence.* During the fifth activity, the categorized evidence put into order of importance for making a diagnosis. The learner considers the prioritized

evidence in terms of importance for diagnosing the case. S/he then moves the individual pieces of information from the categorized list to the new prioritized list, in descending order of importance. Any information not placed in the new list is subsequently discarded. When the sorting is finished, the learner clicks on the *Next* button.

6. *Expert Trace.* During the final activity, the learner examins an Expert's Prioritized List. S/he then has the opportunity to compare this list to their own. This activity concludes each BioWorld case.

Initial Modifications

Based on the general characteristics of the new populations, certain initial modifications were made to the system. It should be stressed that these modifications do not affect the overall learning model of BioWorld. As an example, it was evident that the cases used in the study should be examined by a medical expert and revised as required to ensure that they were technically sound and presented at a level of difficulty that would be suitable for this population.

Modifications to BioWorld for the current study were:

- *Allowing the user to proceed with post-problem reflection with or without a correct diagnosis.* This modification was made to 1) make the activity resemble a real diagnostic situation (e.g., when diagnosing a case in a medical setting, practitioners are not prompted repeatedly until they make the correct diagnosis), and 2) ensure that participants would complete each case including the post-problem reflection activities.
- *Removing the Hypothesis List.* The BioWorld system provides a static Hypothesis List that is available throughout the Problem Solving Phase. Each BioWorld case is on that list. The list was removed since it would have constrained the number of diagnostic categories and therefore the problem space, thereby cueing participants. Removing this list also allowed participants to post multiple differential diagnoses, and to list hypotheses and final diagnoses at varying levels of abstraction (e.g., from general such as "infection", to specific such as "Shigella").
- *Activity 4: Categorizing.* The activity of Categorizing (reviewing data and categorizing it into types such as "Case History") was left in, but it was not essential that participants complete it if they found it simplistic or not helpful. (Briefly, this activity is more suited to the original BioWorld population, high school students, who are learning at a more basic level.)

At the time of the study, it was not possible to implement these refinements to the within the BioWorld software itself. Instead, paper-based versions of the cases were developed that incorporated these refinements. All of the activities and resources were presented in the same order as in BioWorld, with the exception of the modifications listed above. In addition, minor modifications for the paper-based cases were also made.

Briefly, the *Evidence Palette* was replicated by having participants use a highlighter to physically highlight important information. During post-problem reflection, participants were directed to review the information they highlighted in order to complete the categorizing activity. During diagnostic testing, the simulated test results normally provided by BioWorld were instead provided by the researcher. Participants were directed to record the results for the tests they ordered on a form designed for this purpose. Finally, instead of the *Select Hypothesis* and *Belief Meter*, participants were prompted by the researcher to provide their hypotheses/diagnoses after each problem solving activity (frequently they spontaneously provided this information). After several participants indicated that they had difficulty specifying a level of certainty in their diagnoses, the researcher no longer included this component and subsequently left it out of the analyses. See Appendix C for a sample of the case materials.

Other Materials

The other materials consisted of: a) a Certificate of Ethical Acceptability (Appendix D), b) a Pre-Questionnaire (Appendix A), c) a set of Instructions to Participants that described the think-aloud procedure (Appendix E), d) a printout of BioWorld's online library (for an example see Appendix F), and e) a Post-Questionnaire (Appendix G). The purpose of the Pre-Questionnaire was to assess participants' experience with diagnosing digestive cases. The Post-Questionnaire elicited participants' opinions concerning the activities they had engaged in with BioWorld.

Validity of the Experimental Task

Prior to conducting the study, one of the experts (Expert 2) who volunteered to provide expertise for study design, was consulted in order to ensure that the experimental task was valid. That is, he was asked to ensure that the resources (e.g., diagnostic tests) available to solve each case "in the real world" were available in the BioWorld cases. The expert examined the cases and suggested modifications to make each case 1) internally consistent (appropriate symptoms, test results, etc. for a disease type), 2) complete in terms of relevant tests, and 3) somewhat challenging for medical students to solve. First, additional diagnostic tests were added to ensure that all appropriate tests that the participants could reasonably be expected to request were available in the case. It was resolved that any other requests would be answered with "results unavailable" based on the understanding that this "incomplete" knowledge situation is frequently encountered in medical practice. As it turned out, this occurred in few instances. Finally, the list of diseases in BioWorld was eliminated since it would artificially restrict the problem space. The researcher then modified each case as directed and presented them to Expert 2 for final approval. These modifications were made prior to the ratings for case difficulty.

Research Design

Table 3 presents the between-subjects factorial research design for this study. A single experimental condition was used. Two groups of 10 participants (N=20) constituted the between-subjects factor. The first group consisted of third-year medical students and the other of residents. The dependent variables consist of: 1) (Basic) time on task, number and length of utterances, and diagnostic accuracy, 2) (Problem Solving) the use of different operators and states (frequency and percentage of overall use, use in relation to other states and operators, use in different problem solving phases), and 3) (Post-Problem Reflection) Prioritized List and Expert Trace items. The performance of students and residents was examined in light of a model of expert problem solving. Because of the limited number of experts, they were not included as a third group in the comparative analysis. The opinions of participants concerning their experience with the BioWorld cases was also assessed and compared using their ratings on the Post-Questionnaire items.

Table 3.

Research Design

Participant Groups	Experimental Conditions
3rd-year Medical Students	S1...S10
Residents	R1...R10
Experts*	E3 & E4

*Note: The analyses focused primarily on the students and residents. The experts' performance provided a model for comparison purposes.

Experimental Procedure

The experimental sessions were conducted on an individual basis by the researcher. They took place in classrooms at two Montreal-area university teaching hospitals. At the outset of each session, the researcher briefly explained the nature of the study to participants and gave them the opportunity to ask any questions. Participants were then presented with a set of the research materials compiled in the following order. The first document was the Consent Form that the participants were asked to read and sign. The next item was the Pre-Questionnaire concerning their experience with medical diagnosis in digestive system disease (Appendix A). Next, participants were asked to read the Instructions to Participants that describe the think-aloud procedure (Appendix E). They were then invited to ask any questions or request clarification concerning any aspect of the study. At this point, the researcher explained that the library (Appendix F) was available at any time if required.

Participants were then asked to attempt to solve the practice case, followed by the three remaining cases. The order of presentation was randomized to avoid order effects. Participants were asked to think aloud while solving the cases. The researcher did not interfere during this process, except in the following ways. First, at certain points (after reading the case history and later after ordering diagnostic tests) the researcher asked participants to specify a diagnosis or differential diagnoses and if possible to indicate their level of certainty in that diagnosis. The researcher also occasionally prompted participants to continue verbalizing if necessary. The experimental sessions were audio- and video-taped.

Written instructions in the case materials directed the participants to complete each case, starting with the problem solving phase. The first activity in this phase was to read a text-based case history, to highlight relevant information with a pen or highlighter, and to provide any diagnostic hypotheses. The second activity was to review patient chart (physical examination) information including vital signs, etc., and to provide any diagnostic hypotheses. The third activity was ordering diagnostic tests. For each test ordered, the researcher verbally provided the simulated result, and the participants recorded the information in the space provided on the case materials. They were also asked to provide diagnostic hypotheses after reading the case history and to provide their final diagnosis after completing diagnostic testing. For the post-problem reflection phase, participants identified the important information they used in making their diagnosis, listed it in order of importance, and compared their prioritized list with that of an expert. The participants were not told the correct diagnosis for any of the cases. They were also not given any information concerning the nature of the cases and the researcher made no evaluative comments concerning their performance. Once s/he had completed the cases, each participant was asked to complete the Post-Questionnaire (Appendix G). The audio-taped think-aloud protocols were subsequently transcribed and transferred into an N-Vivo project (See *Data Handling: QSR N-Vivo*).

Pilot Study

A pilot study was conducted at one of the Montreal hospitals. Two medical students were recruited and asked to individually complete the experimental procedure as outlined above. No alterations were made in the materials or the procedure based on the results of the pilot study. As such, the protocols from these participants were used as data in this study (participants S1 and S2).

Data Analysis

The research questions and the analyses conducted in order to answer them are listed in the following section.

Data Handling: QSR N-Vivo

QSR-N-Vivo is a software application designed for qualitative research purposes, typically analyzing verbal data. N-Vivo was used to set up a project for this study including building and modifying the coding scheme, importing data (protocols) into the project, applying the codes to the data, and conducting a number of search types on the coded protocols. Each of the 22 original think-aloud protocols (1 per participant) were divided into three to create one protocol per case. An N-Vivo project was created, and the 66 new protocols were subsequently imported into the project and used in the data coding and subsequent analyses.

Research Question #1: General Analyses

1. ***How does the general performance (time on task, protocol length, diagnostic accuracy) of students and residents differ as they attempt to diagnose internal medicine cases based on case history, physical examination, and diagnostic testing activities? How does their performance compare with that of experts?***

Mean Time to Solve a Case

The time required for each participant to solve each case was measured based on the videotaped transcripts. Basic descriptive statistics (mean time per group, range, and standard deviations) were also calculated. A two-factor (group and case) Analysis of Variance (ANOVA) with repeated measures on one factor (case) was conducted to examine the differences between the time required for students and residents to solve each case. A significant result was found, and a post-hoc Tukey test was performed to determine which differences were significant. It is important to note that the case presentation was counter-balanced; the individual participants did not complete the cases in the same order.

Protocol Length, Number of Utterances, & Average Length of Utterance

In order to provide a basic description of the data, an N-Vivo *Document Search* was conducted. The results included the number of utterances per individual based on paragraph returns in the N-Vivo documents. The descriptive statistics calculated include the mean number of utterances, the average length of utterance per group, and the standard deviations. The mean protocol length (number of utterances) and the mean length of utterances were also tabulated using *Document Search* in NVivo. The number of section headings and the number of utterances made by the researcher (in response to requests for data, test results, etc.) were also subtracted from the results.

A two-factor (group and case) Analysis of Variance (ANOVA) with repeated measures on one factor (case) was conducted to examine the differences in the number of utterances made by the students and residents for each case. A post-hoc Tukey test was again performed to determine which differences were significant.

Diagnostic Accuracy

Diagnostic accuracy was assessed by expert ratings on a 5-point Likert scale (1 = wrong, 2 = poor, 3 = neutral, 4 = good, and 5 = excellent). Expert 2 was presented with a list of all of the participants' final diagnoses and hypotheses and asked to rate the accuracy of each given the information available in the case. The expert recorded a rating in the space provided beside each diagnosis on the list. The list contained only the case number and name, and the diagnoses and hypotheses themselves. No information about which participants had offered them was included. The order of the list items was also mixed to avoid order effects. In several instances, students did not provide a final diagnosis, and ratings were not assigned. No information about participants was given to the expert, and the order of responses was randomized. Mean accuracy ratings were calculated by group and case. A Chi-Square analysis of the frequency of ratings for students and residents across cases was also conducted.

Research Question #2: Problem Solving

2. ***How does the problem solving of students and residents differ as they attempt to diagnose internal medicine cases based on case history, physical examination, and diagnostic testing activities? How does their problem solving compare with that of experts?***

A number of analyses are reported in this section in order to address this research question. The first analyses presented are the expert models, which provide an example of competent performance on the three cases. These models are used as a context for examining the problem solving of the students and residents. The models are referred to throughout the analysis of problem solving performance.

Expert Models

Expert models of problem solving (one per case) were created based on the protocols of experts. Each model summarizes the movement of the two experts through the problem space for one case, in terms of three main aspects that were identified during the course of the analysis: plans, data, and hypotheses. The models were developed by examining the protocols of both experts from the problem solving activities for each of the cases. These models characterize experts' problem solving, and are useful for highlighting some of the differences between experts, students, and residents. Developing the models involved working through the individual expert protocols for each case to identify the:

1. raw data (Case History and Patient Chart) that was important,
2. the diagnostic tests that were ordered and how the results were interpreted,
3. hypotheses that were generated at various points during problem solving,

4. impact of new data on existing hypotheses,
5. kinds of goal statements made,
6. actions that were carried out to achieve stated goals, and
7. final diagnoses provided.

Once these factors were identified, diagrams were created for the models that depict the experts' use of data, hypotheses, and plans in conjunction with the learning activities (Case History, Physical Exam, Diagnostic Testing) in which these factors occur. The order in which the data, hypotheses, and plans were mentioned is recorded in the models by the numbered circles beside each item. The models, therefore, provide a trace of the knowledge used and actions taken to solve each case. Finally, bold text is used in the models to show steps that were common to both of the experts, while normal text shows steps taken by only one of the experts. For more details about the layout of the models, see *Chapter 4, Expert Models: Organization of the Models.* One student model was also constructed following the same procedure as the expert model and used for comparison purposes.

The Coding Scheme

In order to examine the participants' cognitive processing, a coding scheme was developed based on relevant research (e.g., Bracewell & Breuleux, 1994; Hassebrock & Prietula, 1992). It was applied to the data and revised following an iterative process that resulted in the development of a set of codes that were a) manageable, and b) detailed enough to distinguish between meaningful cognitive factors. Following Information Processing Theory (IPT), two basic categories in the coding scheme were cognitive operations and knowledge states. The separation of these two categories allowed for a fewer number of codes overall and flexible pairing of operators with knowledge states.

The coding scheme was applied to a set of verbal protocols (approximately 30). Further specification and modification resulted from repeatedly coding samples of the data. Analysis of the data was particularly helpful in specifying knowledge states related to strategic thinking (e.g., planning), as well as other states and operators. The coding scheme was applied to half of the protocols and modified several times based on the data and the results of the task analysis.

The coding scheme was constructed in N-Vivo in a hierarchical tree structure with 4 high-level nodes: 1) Activities, 2) Operators, 3) Knowledge States, and 4) Other. These nodes represent both task-specific elements and more general components of thought. That is, the first node, *Activities*, includes 6 sub-codes, one for each BioWorld activity. The next two nodes correspond to two key elements of Information Processing Theory, *Operators* and *Knowledge States*. Finally, the fourth node, *Other*, includes a set of domain-independent codes that were not included in the analysis because they do not directly contribute to answering the specified research questions.

All of the protocols were coded using N-Vivo. The coded protocols were also organized to allow for conducting a number of different search types. For example, the individual protocols were assigned "Attributes" for participant type (student, resident, and expert) and for case (Celiac, Hepatitis A, Shigellosis). Each protocol was also coded for a BioWorld activity. Briefly, this involved applying one activity code (e.g., "Case History") to the entire section of each protocol that dealt with that activity. The search types included Boolean searches to determine the co-occurrence of operators and knowledge states, and to obtain a breakdown of operator use by group. The results of the various searches conducted and subsequent analyses are reported in this section.

The final coding scheme, including the N-Vivo code ID numbers, code names, definitions, and examples, is presented in Table 4. The ID numbers denote positions in the coding hierarchy. That is, IDs with a single digit represent the four high-level nodes in the coding hierarchy. Sub-codes under these are automatically numbered by N-Vivo, and will have two or more numbers, separated by a space, depending how many code levels they are nested within. For example, the ID number 1 1 3, would identify a coding address that falls under the first high-level node (1), and under the first second-level node (1) in the third position (3).

The final coding scheme was subsequently used to code all of the protocols in N-Vivo and to run a number of searches on the coded documents. All documents were coded for BioWorld phase, operations and knowledge states as specified in the coding scheme. Frequency and percent counts were calculated for the number and percentage of utterances for each BioWorld activity, and for problem solving versus post-problem reflection activities. The BioWorld phase codes were examined by comparing the number of utterances per BioWorld phase across cases and groups. Frequency and percentage use of all operators and knowledge states were also tabulated.

Table 4.

N-Vivo Code ID#s and Names, Definitions, and Examples[1]

Code ID# & Name	Definition	Example
(1) Activity*	**BioWorld activities**	
(1 1) case history	verbalizations concerning case history	*"So, she recently returned from Mexico."*
(1 2) physical exam	verbalizations during physical exam phase	*"Her temperature is 38.5 degrees."*
(1 3) testing	verbalizations during testing phase	"Give me a hemoglobin."
(1 4) categorizing	verbalizations during categorizing phase	*"For symptoms there's nausea, vomitting…"*
(1 5) prioritizing	verbalizations during prioritizing phase	*"Number 1 for me was positive Hep A (titer)."*
(1 6) expert trace	verbalizations during expert trace phase	*"The expert put that first…"*
(2) Operators*	**cognitive actions performed on knowledge states**	
(2 1) state	a statement or assertion, excluding other operators	*"He has intermittent diarrhea…"*
(2 2) request	make a request (e.g., diagnostic test)	*"Give me the bilirubin (test result)."*
(2 3) explain	offer an explanation for any aspect of case, a statement, or an action	*"I'm doing this to see how badly dehydrated he is."*
(2 4) evaluate	evaluate knowledge state in terms of utility or quality	*"The jaundice just confirms it."*
(2 5) link	suggest a link or connection between knowledge states	*"The white count supports an infectious cause."*
(2 6) read	read verbatim from case materials (e.g., Case History)	
(3) States*	**knowledge states**	
(3 1) Data*	**types of raw data**	
(3 1 1) case info	patient information (symptoms, vital signs, etc.)	*"She has a fever, nausea and vomiting."*
(3 1 2) test result-s	values returned from a requested diagnostic test	*"The folate level is low."*
(3 1 3) test	a diagnostic test	*"Small bowel follow through (SBFT)."*
(3 1 4) missing	knowledge state that is unavailable	*"We don't know what else her doctor ordered at that time."*
(3 1 5) library	information contained in on-line library	*"(The library) says the incubation period is…"*

[1] Note that some codes were not applied to the data, because they are high or intermediate level-codes that are further broken down into sub-codes. These codes are marked with (*) and examples are not provided. The only exception is the code Plan, which was coded, along with sub-codes.

Table 4, continued

(3 2) Inferred*	**knowledge states derived from raw data**	
(3 2 1) patient condition	patient's inferred physiological condition, not raw data	*"The patient has some dehydration."*
(3 2 2) diagnostic solutions	hypotheses/diagnoses, at varying levels of specificity	*"It is an infection."/"She has Hepatitis A."*
(3 2 2 1) hyp-new	tentative diagnosis for a case, not offered previously	*"My first thought is Irritable Bowel."*
(3 2 2 2) hyp	tentative diagnostic hypothesis, offered previously	*"So it is back to Ulcerative Colitis."*
(3 2 2 3) final diag	diagnostic hypothes(e)s offered as the final solution	*"It seems to be Irritable Bowel."*
(3 2 3) strategic*	goals and plans for problem solving	
(3 2 3 1) plan	Set of 2 or more goal statements, marked by intentionality & future/potential action	*"First I will do the physical exam. And next I will order a CBC."*
(3 2 3 1 1) action	action marked as current or next	*"I will now examine liver function."*
(3 2 3 1 2) potential action	possible action, marked by qualifier, ability	*"I might subsequently take a culture."*
(3 2 3 1 3) treatment	a treatment goal for a patient	*"I will replete his volume."*
(3 2 4) self*	knowledge states concerning one's own characteristics	
(3 2 4 1) knowledge	own knowledge	*"I'm not familiar with Reye's Syndrome."*
(3 2 4 2) performance	own performance or actions (summative)	*"I forgot to get an HIV (test result)."*
(3 2 4 3) regulation	own progress *during* problem solving	*"What else should I do?"*
(3 2 5) relevant	information presented as potentially relevant to the case	*"Diarrhea can be chronic or acute."*
(4) Other*	**Domain-independent coding**	
(4 1) determiners & quantifiers*	**words/phrases that identify mathematical relationship between knowledge states (number, degree, and order)**	
(4 1 1) number	Word/phrase denoting specific quantity of concepts	*"Three problems must be dealt with."*
(4 1 2) degree	Word/phrase denoting a non-specific amount of an attribute	*"The level is slightly elevated."*
(4 1 3) order:	indicates an order or ranking in a list	*"This would be as a last recourse."*
(4 2) truth value & modals*	**words/phrases that add meaning concerning the truth value or possibility of a statement**	
(4 2 1) negative	indicates a negative truth value	*"This is not cancer."*
(4 2 2) qualify	qualifies a statement	*"This is probably not cancer."*
(4 2 3) necessity	unstated conditions exist that make a concept a necessity	*"This fluid loss must be corrected."*
(4 2 4) ability	unstated conditions exist that will lead to a concept	*"The fluid loss can be corrected."*

Other analyses were also conducted using the *Document Search* and Boolean searches (*Matrix Intersection*) function in N-Vivo. Included were the co-occurrence of each of the *Link, Explain* and *Evaluate* operators with each knowledge state, by case and group. The occurrence of new hypotheses by BioWorld activity for each case and group was also tabulated using this search option. A two-factor (group and activity) Analysis of Variance (ANOVA) with repeated measures on one factor (activity) was used to determine whether there was a significant difference between the groups in the number of new hypotheses generated by BioWorld activity. A post-hoc Tukey test was also performed to determine where the significant difference occurred.

The use of strategic knowledge (plans, etc.) was further examined. All utterances coded as plans (including the sub-codes for goals) were extracted using an N-Vivo search and re-examined. The number of steps per plan by case and group were tabulated. In addition, the content of plans by group were examined and a small set of codes for 5 different content types was developed. All of the participants' plans were coded as one of these types:

1. *Testing*. Lists a series of intended actions for diagnostic testing. For all groups, more participants generated plans related to Testing than to other content areas.
2. *Treatment*. Identifies a set of actions that address treatment of the patient's condition. One student, 5 residents, and one expert specified this type of plan.
3. *Gather information*. Specifies types of additional information that s/he would obtain about a patient (e.g., specific case history information). One student, S8, generated 5 plans concerning information s/he wanted to obtain from further case history investigation. One other student, one resident, and both experts also listed plans for gathering additional information.
4. *Diagnostic list*. Identifies diagnostic hypotheses under consideration in relation to actions.
5. *General*. Lists or general strategies or steps being undertaken.

The results of this coding were tabulated by individual, group, and case, and are presented in the *Results* section.

Inter-Rater Reliability

A postdoctoral fellow at the University of Pittsburgh coded a subset of the protocols (3 protocols: one each from a student, a resident and an expert from different cases, totaling over 400 utterances). An initial inter-rater reliability of approximately 85% was established. Further discussion between the researcher and the postdoctoral fellow resulted in the collapsing of some coding categories (e.g., three subcategories of the operator code "Link" were collapsed). The inter-rater reliability was then established at over 94%.

Research Question #3: Post-Problem Reflection

3. ***How does the post-problem reflection (categorizing data, prioritizing data, and comparing to an expert list) of students and residents differ? How do their reflective actions compare with those of experts?***

Participants' written responses for the *Post-Problem Reflection* activities (Categorizing, Prioritizing, and Expert Trace) were examined. The information in the prioritized lists was labeled as either case information (history and physical examination), or diagnostic testing data, and the number of items of each type were tabulated by case and group. The data (e.g., case history or testing) and its ranking in the participants' prioritized lists were also tabulated by group. The participants' responses to the Expert Trace were broken down by response type, and their comments were further categorized into three types. Some participants' comments are summarized and examples are provided in the next chapter.

Additional Findings: Participant Ratings

Post-Questionnaire

Finally, the Post-Questionnaire results (5-point Likert rating scale) were examined. Basic descriptive statistics (means and standard deviations) were calculated in order to summarize the opinions of the groups concerning BioWorld. The results of the Post-Questionnaire are also compared with the expert ratings for diagnostic accuracy.

Chapter Summary

This chapter has outlined the methodology for the study and provided an overview of the BioWorld learning environment including modifications made for the current study. The data and analyses to be used to answer each of the specified research questions have also been specified. The following chapter will present the *Results and Discussion.*

CHAPTER 4: RESULTS AND DISCUSSION

Introduction

Throughout this chapter, the study results are listed and discussed in relation to the relevant medical expertise literature. The results are organized according to the research questions presented in Chapters 1 and 3. Each of these sections ends with a summary of the results in relation to the research question it addresses. The first section examines the general performance of participants including several data characteristics (e.g., mean time to solve each case, the length of verbal protocols, number of utterances per protocol), and the diagnostic accuracy ratings. The second section presents the results relating to problem solving activities including the expert models and the application of the coding scheme. The third section briefly examines the post-problem reflection activities (Categorizing, Prioritizing, and Expert Trace). Finally, the last section presents additional findings, specifically the analysis of the Post-Questionnaire data. The following chapter draws on these results to make a set of recommendations for the future development of BioWorld for medical education settings.

General Results

Answering Research Question #1

1. ***How does the general performance (time on task, protocol length, diagnostic accuracy) of students and residents differ as they attempt to diagnose internal medicine cases based on case history, physical examination, and diagnostic testing activities? How does their performance compare with that of experts?***

Data Characteristics

Mean Time to Solve Each Case

The mean time for the two groups to solve each case is presented in Table 5. The time experts took to solve cases is also reported although statistical comparisons have not been made. The mean time includes both the problem solving and post-problem reflection activities. There was a considerable variation in the time spent by individual participants in both groups to complete each case, as indicated by the range and standard deviation values. For both groups there is a linear decrease in mean time with decreasing case difficulty; however, the differences between Case 1 and 2 are relatively small, especially considering the standard deviations. Case difficulty apparently had an impact on solution time for each group, with considerably less time required to solve Case 3 (easy) than either of the other cases. All groups also showed the largest discrepancy between the time to solve this case and the time to solve the other two cases. The ranges and standard deviations also show that the participants in each group varied more in the time it took to complete these cases as compared to Case 3. Finally, the overall means by group show that residents spent the least time overall, followed by students and then experts. The overall means by case show a (approximate) linear decrease overall with decreasing difficulty.

Table 5.

Mean Time to Complete Each Case by Group

	Case 1: Celiac (difficult)	Case 2: Shigellosis (medium)	Case 3: Hepatitis A (easy)	Overall Mean (by group)
Students (n=10)				
Mean Time (min:sec)	14:44	14:31	10:23	39:38
Mean Time (sec)	884	871	623	
Range (min:sec)	(8:05 – 21:05)	(10:37 – 20:37)	(7:58 – 14:14)	
SD	3:54	4:02	2:36	
Residents (n=10)				
Mean Time (min:sec)	14:23	13:03	7:58	35:24
Mean Time (sec)	863	783	478	
Range (min:sec)	(7:40 – 21:37)	(6:50 – 17:43)	(6:08 – 9:59)	
SD	4:08	2:51	1:08	
Experts (n=2)				
Mean Time (min:sec)	20:17	16:24	9:05	45:46
Mean Time (sec)	1217	984	545	
Range (min:sec)	(18:25–22:10)	(13:54-18:55)	(8:28 – 9:42)	
SD	2:39	3:32	0:52	
Overall Mean (by case)	49:24	43:58	27:26	

A two-factor (group and case) Analysis of Variance (ANOVA) with repeated measures on one factor (case) determined that there was no significant difference in solution time between the groups, but there was a significant difference in the time required to complete the cases [$F(2, 36) = 23.03$, $p = .0001$]. The post-hoc Tukey test was performed to determine which cases differed (HSD = 126.9)[2]. A significant difference was found between Cases 1 and 3 (332.45) and Cases 2 and 3 (276.75). For Case 3, both students and residents spent considerably less time than they did on the other cases (approximately 5 minutes). There was no significant difference between Cases 1 and 2 (45.7). There was also no interaction between the groups and cases. It should be noted that the case presentation was counter-balanced: the participants did not solve the cases in the same order.

Typically, studies have found that there is a linear decrease in both case solution time and case reading time with increasing expertise (e.g., Rikers et al., 2000a, 2000b). However, as shown in Table 5, the experts spent more time on Case1 (20:17) and 2 (16:24) than either of the other groups, a finding that probably reflects lengthier or more elaborate processing or explanations for various aspects of these more difficult cases. Several subsequent analyses support this conclusion. It is also possible that the nature of the experimental task (e.g., the non-interactive case history and physical examination) contribute to this result. That is, they were not able to request specific case history or physical examination data that would probably have helped to distinguish between some diagnostic hypotheses. As a result, they may have engaged in more processing as they tried to consider and maintain all of the possible hypotheses until they could use diagnostic tests to start narrowing them down.

Both students and residents showed a linear decrease in mean time to solve the cases with decreasing case difficulty. For Case 3 (the easy case), experts spent more time than residents, but less time than students; perhaps, the straightforward nature of the case led to less elaboration as compared to the other cases. Experts also showed a linear decrease in the mean time to solve the cases with decreasing case difficulty. While experts usually take less time than less-experienced participants (e.g., Custers et al., 1996), other researchers have also found that with difficult cases, experts sometimes take longer to complete parts of the diagnostic task than non-experts. In the Norman et al. (1994) study, experts took longer than three other groups (residents with different levels of expertise) to review case information for difficult nephrology cases. The researchers attributed that finding to experts' more elaborate explanations. This is also a plausible explanation for the results of this study. It is further supported by findings reported in the following section. A contributing factor to the lengthier expert protocols, as further

[2] An alpha level of .05 was used for the post-hoc analysis. The df_{error} used was 36, however 30 was used because it was the next smallest value in the table (Glass & Hopkins, 1996).

analyses will demonstrate, is the experts' considerably lengthier processing of case history information. Finally, the experts spent a lot less time on the easy case than on the other cases. It makes sense that they do not elaborate to the same extent when a case appears to have a clear-cut solution.

Protocol Length

Mean number of utterances. Table 6 presents the mean number of utterances and standard deviations by group and case. Experts are also included in the table for reference purposes. A two-factor (group and case) Analysis of Variance (ANOVA) with repeated measures on one factor (case) was conducted to examine the differences between the mean number of utterances for students and residents by case.

Table 6.

Protocol Length: Mean Number of Utterances

	Case 1: Celiac (difficult)	Case 2 : Shigellosis (medium)	Case 3: Hepatitis A (easy)	Overall Mean (by group)
Students (n=10)				
Mean	139.2	124.7	94.6	358.5
SD	12.29	9.42	9.42	
Residents (n=10)				
Mean	125.7	103	79.5	308.2
SD	8.61	8.76	6.88	
Experts (n=2)				
Mean	289.5	334	134	757.5
SD	19.8	157	16.3	
Overall Mean (by case)	146.73	133.86	91.32	

This analysis determined that there was no significant difference between the groups, but there was a significant difference in the number of utterances by case [F(2, 36) = 18.69, p =.0001]. A post-hoc Tukey test was performed to determine which cases differed (HSD[3] = 18.4). Significant differences were found for all pair-wise comparisons. In other words, the groups spent significantly less time with decreasing case difficulty. Also, there was also no interaction between the groups and cases. Again, it is important to note that the case presentation was counter-balanced: the individual participants did not solve the cases in the same order.

[3] An alpha level of .05 was used for the post-hoc analysis. The df_{error} used was 36, however 30 was used because it was the next smallest value in the table (Glass & Hopkins, 1996).

The experts made more than twice as many utterances than the other groups overall (overall means, far right column), and they had a much higher mean number of utterances for each case than either of the other two groups. Students made more utterances than the residents by case, but the differences are not so great given the standard deviations. There was a large discrepancy between the number of utterances made by the two experts on Case 2, as indicated by the standard deviation. The overall means by group show that residents made fewer utterances overall, followed by students and then experts. This also supports the finding that experts engaged in more elaborate processing than the other groups.

The overall means by case show a (approximate) linear decrease overall with decreasing case difficulty. Experts had much lengthier protocols for the two more difficult cases than for the easy case, and their protocols were much lengthier than students and residents.

Mean length of utterances. The mean number of words per utterance by group and case are presented in Table 7. The results for students and residents were very similar both by case and overall. The standard deviations for the students and residents were quite similar across the cases, ranging from 0.99 to 1.38. These analyses are important for confirming that the number of utterances (see Table 6) actually represent a real difference between the groups, rather than instead being a product of varying utterance length.

Table 7.

Words Per Utterance by Group and Case

	Case 1: Celiac (difficult)	Case 2: Shigellosis (medium)	Case 3: Hepatitis A (easy)
Students (n=10)			
Mean	6.23	6.72	5.97
SD	1.03	0.99	1.02
Residents (n=10)			
Mean	6.68	6.43	5.73
SD	1.38	1.19	1.12
Experts (n=2)			
Mean	9.02	10.43	9.79
SD	0.5	1.81	1.12

Finally, the expert utterances were consistently somewhat longer on average both by case and overall. This may be attributable to experts' greater use of certain types of utterances (e.g., different operators), a suggestion that is supported by subsequent analyses (see the analyses conducted in response to research question #2, *Problem Solving*). There was a large discrepancy between the two experts on Case 2 only.

Diagnostic Accuracy

Expert ratings of the diagnostic accuracy of each of the participants' final diagnoses are presented in this section. Table 8 lists the cases, the actual diseases, and the correct diagnoses based on the available data. Expert 2 provided the final diagnoses and additional acceptable hypotheses for each case, as presented in Table 8. The cases differed in the level of specificity possible based on the available data. These differences are reflected in the information in the table.

Table 8.

Case #, Actual Disease, and the "Correct" Diagnosis Based on the Data

Case	Diagnosis Based on Available Data
1: Celiac	Malabsorption, with Celiac disease a possibility but not confirmed
2: Shigellosis	Infection (bacterial) with Shigella, E-coli, Salmonella, as possibilities, not confirmed
3: Hepatitis A	Hepatitis A, confirmed by positive test result for Hepatitis A titer

Expert 2 also provided the diagnostic accuracy ratings for all final diagnoses. A summary of the results is provided in Table 9. For Case 1 (Celiac), students had a lower mean accuracy rating (1.29) than residents (1.7). Three students did not provide a diagnosis for this case. Both experts also considered Celiac disease and ordered the appropriate diagnostic test (Anti-Gliadin Antibodies). While they did not specifically state *Celiac Disease* in their final diagnoses, they had both previously discussed it as a possible cause of the malabsorption. Two residents actually did state a correct basic diagnosis (*Malabsorption*) and received a rating of 4. The two experts both provided the correct diagnosis also, and received the same rating. No students correctly diagnosed this case.

For Case 2, Shigellosis (moderately difficult), the mean ratings for students and residents were higher than for Case 1 (students = 3, and residents = 2.3). However, two students failed to provide a diagnosis and none of them provided the correct diagnosis. In this instance, both experts did mention the "correct" diagnosis during the course of their problem solving, but Expert 3 did not specifically mention the potential causes of the infection as part of her final diagnosis and received a slightly lower rating.

Table 9.

Expert Ratings for Diagnostic Accuracy (Final Diagnoses)

Group	Case 1: Celiac (difficult)	Case 2: Shigellosis (medium)	Case 3: Hepatitis A (easy)	Overall Means
Students (n=10)				
Mean	1.29	2.4	5	2.77
SD	0.49	0.93	0	
Residents (n=10)				
Mean	1.7	2.3	4.8	2.93
SD	1.25	0.48	0.63	
Experts(n=2)				
Mean	4	3.5	5	4.17
SD	0	.71	0	

(1 = wrong, 2 = poor, 3 = neutral, 4 = good, and 5 = excellent)

Finally, for Case 3 (Hepatitis A), all participants except one resident gave a correct diagnosis. As mentioned, this was a relatively easy case, and the diagnosis could be confirmed by a specific test (Hepatitis A titer). The resident did mention Hepatitis A in his initial discussion of the case, but neglected to actually order the Hep A titer during diagnostic testing. He noticed the oversight after viewing the Expert Trace.

A Chi-Square analysis of the frequency of the ratings for students and residents across cases, produced a value of $\chi^2 = 0.47$ (df = 4, p <.05), which is not statistically significant. A t-test was also performed for Case 1, which was also not significant (t = -0.82, df = 15, $\alpha = 0.05$, p = 0.42). This indicates that the students and residents did not differ in diagnostic accuracy across the cases. The differences in variance and the small amount of data contribute to the non-significant results.

The expert ratings for diagnostic accuracy follow the expected general trend of increasing diagnostic accuracy with level of expertise. Also, for the students and residents, the diagnostic accuracy ratings increased from Case 1 to Case 3. For Case 1 students had a lower mean accuracy rating (0.9) than residents (1.7). On Case 2, students and residents were very similar in diagnostic accuracy (students = 2.4, residents = 2.3). Students, residents, and experts scored essentially at the same high level for the easiest case (Case 3: Hepatitis A). It is also interesting to note that 1) on Case 1, two of the residents received high ratings (4) while none of the students did, and 2) on Case 2, one student also received a high rating (5). The residents' most variable performance is seen on Case 1, as shown by the higher standard deviation.

Numerous studies have shown that diagnostic accuracy increases linearly with level of expertise (e.g., Rikers et al., 2003) and the current findings are basically consistent with this, although a non-significant result was obtained. Students did not perform as well as residents. For Case 1 (difficult), no students correctly diagnosed the case, and three failed to provide a diagnosis. Two residents correctly diagnosed the case. For Case 2 (medium), two students failed to provide a diagnosis, but overall the students received a similar accuracy rating as residents. For Case 3 (easy), diagnostic accuracy ratings were high and basically the same for all groups, and all participants provided a diagnosis. The experts had higher accuracy ratings for each case than either of the other groups.

Summary

Differences in case solution time and the number of utterances across levels of expertise were not significant. While the differences between the groups are in the expected direction (i.e., residents on average took less time and made fewer utterances than students), there was fairly large within-group variation. Also, there was a significant decrease in both the number of utterances and solution time with decreasing case difficulty. Interestingly, unlike the findings from several studies that report that experts are faster at solving cases, this study found that they took longer, made many more utterances, and made somewhat longer utterances than non-experts. These results suggest that experts engaged in more processing, the nature of which will be examined more closely in the following section (*Problem ving*). Finally, students and residents did not differ significantly in diagnostic accuracy. However, a large difference was observed between accuracy ratings for the groups across the cases.

Problem Solving

Answering Research Question #2

2. ***How does the problem solving of students and residents differ as they attempt to diagnose internal medicine cases based on case history, physical examination, and diagnostic testing activities? How does their problem solving compare with that of experts?***

This section examines participants' problem solving as they completed the cases. The expert models are presented first, since they provide an example of competent performance on the three internal medicine problems. They further highlight some interesting elements of expert problem solving in this domain, and provide models to compare with the problem solving of the other two groups. In general, based on previous research in medicine and other domains, it might be expected that there would be evidence of a progression from novice to expert in several aspects of performance. This expectation is confirmed with some of the analyses performed in this section. The main focus of the problem solving analysis is on the results of the coding scheme application. This provides an in-depth analysis of the operators and knowledge states used by the groups as they attempted to solve the cases. The use of strategic knowledge (e.g., planning) is also examined in detail.

Expert Models

The main contribution of the expert models is their depiction of the expert's cyclical process of reviewing case information, planning, and testing. These results are briefly discussed in conjunction with other research findings.

Organization of the Models

Figures 2 - 4 present the expert models, which are organized as follows. Each model provides an integration of the steps taken by the two experts as they completed a case. By following the numbered circles, the reader can re-trace the progression of steps made by the experts to complete the cases. The utterances common to both experts appear in each figure in **bold** type. These constitute the main part of the model, indicating information that was used by both experts. Other utterances made by one of the experts only are included in regular type. The top of each figure shows three boxes denoting three types of statements, *Planning/Goal Statements*, *Reviewing/ Collecting Data*, and *Hypotheses*. The BioWorld activities are depicted on the left side of each figure in three boxes labeled *Case History, Patient Chart*, and *Diagnostic Testing*. Each utterance is placed in the figure horizontally according to its type, and vertically according to which BioWorld activity it occurred in. Arrows are used in the figures to link together planning/ goal statements with the actions taken to achieve those goals. Typically, the actions involved ordering a specific test to determine if a diagnostic hypothesis was valid or not.

So, for Figure 2, the box labeled #1 under *Reviewing/Collecting Data* corresponds to some of the experts' initial utterances, which occurred as they read the case history and identified information that was relevant. Box #2, under *Hypotheses*, contains the hypotheses listed by the experts based on their reviews of the case history. The next utterance appears in box #3, where a goal/planning statement is made. Following the numbered boxes, the solution of the case (for experts) then becomes a cycle of these activities, until the final diagnosis is provided. Obviously, not all utterances are represented in the models. Instead, they depict the data-plan-test cycles that were extracted from their protocols.

Characteristics of Expert Problem Solving

Comparing experts. Some striking characteristics of expert problem solving become apparent through constructing the Expert Models. First, as expected, the problem solving of the two experts was not identical. Their thoughts and actions do have many common features, but include unique components as well. So, the models include both their shared thoughts and actions, as well as some of these unique components.

Expert strategic processes. Another characteristic made very clear by the expert models, is that the experts seem to collect information and use it strategically in subsequent problem solving (e.g., use it to advance through the problem space). For example, with each case the experts generate multiple goals or plans related to diagnostic testing, and subsequently follow up on them by ordering tests. The link between plans/goals and the actions carried out to achieve them is made clear in the models. Planning is typically not seen to this extent in the protocols of other participants.

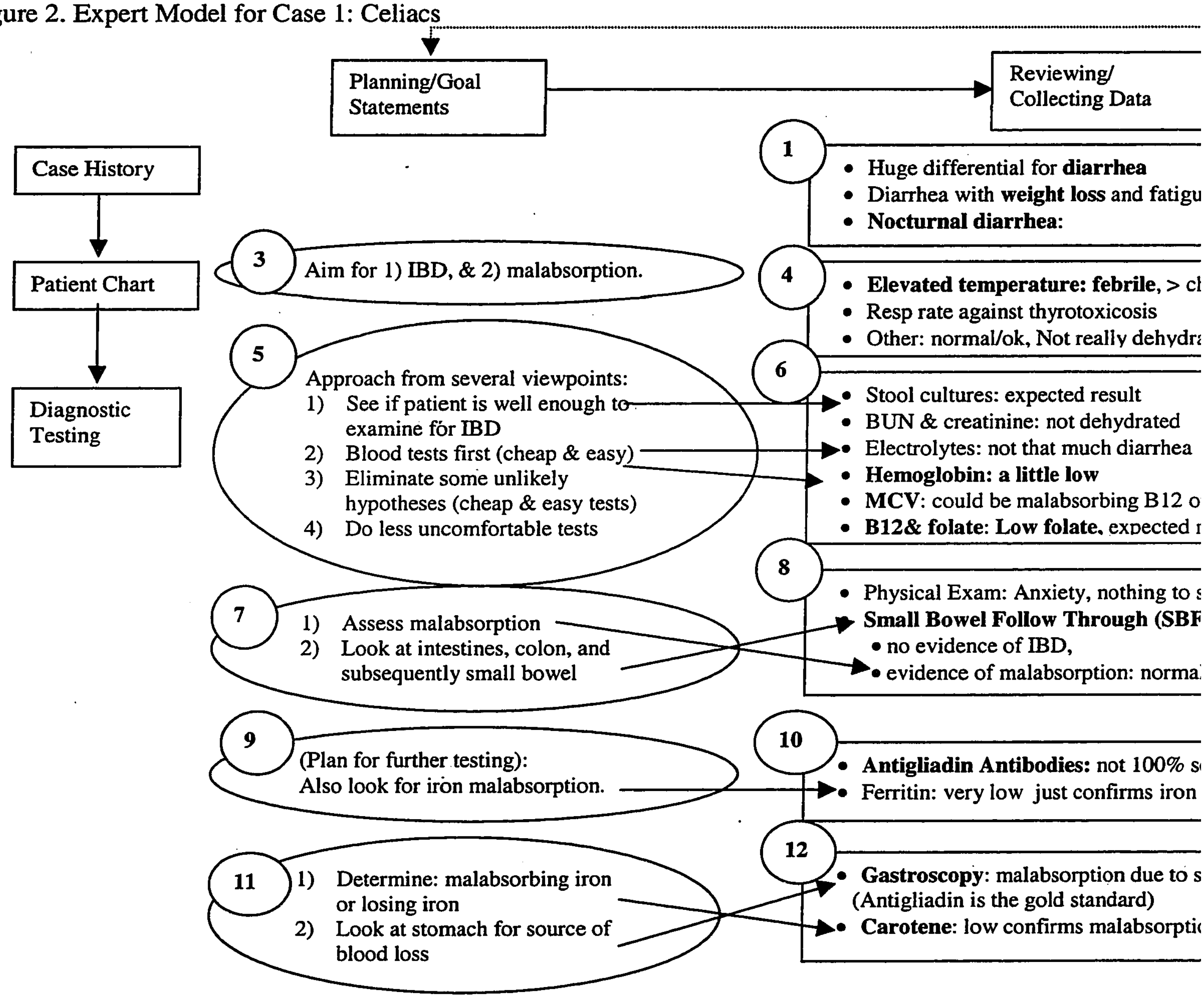
Another feature of expert plans is their variety. They include: 1) very general plans (e.g., to gather information to confirm or eliminate specific hypotheses and conduct tests to look for complications), 2) plans for specific tests (e.g., what to look for on physical examination), 3) plans to search for particular etiologies (e.g., malabsorption), and 4) ordered lists of actions (e.g., to first look for one disease and then for another). Expert use of plans/goals will be discussed in more detail and compared to that of the students and residents later in this chapter.

The expert models are consistent with findings from other studies. Patel et al. (1994) found that experts develop hypotheses quickly in response to data, and they accommodate additional data as it is acquired. The models show that the data are used to identify multiple hypotheses and subsequently new data are used to narrow the hypotheses. The models also give some support for expert problem solving as *data-driven* rather than *hypothesis-driven*. That is, the data collected is reviewed and used to specify hypotheses and subsequent plans and actions. The data is the major focus during problem solving. This finding is supported by a large number of studies (e.g., Allen et al., 1998; Barr & Feigenbaum, 1981; Lesgold, 1988; Norman et al., 1994; Patel et al., 1991; Patel & Groen, 1986; Patel & Ramoni, 1997; Reimann & Chi, 1989). These models depict the purposeful movement from multiple initial hypotheses, to testing, and to the removal of some hypotheses and selection of the correct hypothesis. Similarly, experts in the Allen et al. (1998) study used efficient evidence-gathering strategies and controlled reviewing of the evidence that resulted in refinements to initial hypotheses. Finally, as Evans and Gadd (1989) suggested, planning in diagnostic problem solving contexts is important for resolving uncertainty and interpreting new information. Planning statements are important in linking the initial data and hypotheses with additional data that can be used to discriminate relevant versus irrelevant diagnostic possibilities.

A Sample Student Model

Figure 5 presents a student model for participant S1 on Case 1 (Celiac). It is an example that can represent several instances of both student and resident problem solving. The expert model for the same case is presented in Figure 2. That is, the students and residents often did not engage in the planning activities that are so evident in the Expert Models, but focused mainly on collecting information and interpreting it in terms of diagnostic hypotheses. On some cases, students and residents did engage in planning, which is described in more detail in the *Problem Solving* section of this chapter. However, none of these participants consistently developed multiple plans as the experts did in order to solve the cases.

Figure 2. Expert Model for Case 1: Celiacs



Reviewing/ Collecting Data

Hypotheses

1
- Huge differential for **diarrhea**
- Diarrhea with **weight loss** and fatigue is a bit ominous
- **Nocturnal diarrhea**:

2
1. **Organic** > functional, serious
2. **Inflammatory Bowel Disease (IBD)**
3. Malabsorption, specifically Sprue
4. Low likelihood of **infection** (HIV or Giardia)
5. Dietary indiscretions/changes
6. **Tumors** & vascular problems (unlikely)
7. Amoebic
8. Laxative abuse, bulimia
9. Endocrine problems
10. thyrotoxicosis

4
- **Elevated temperature: febrile,** > chance of HIV/infection
- Resp rate against thyrotoxicosis
- Other: normal/ok, Not really dehydrated

6
- Stool cultures: expected result
- BUN & creatinine: not dehydrated
- Electrolytes: not that much diarrhea
- **Hemoglobin: a little low**
- MCV: could be malabsorbing B12 or folate
- **B12& folate: Low folate**, expected reverse results

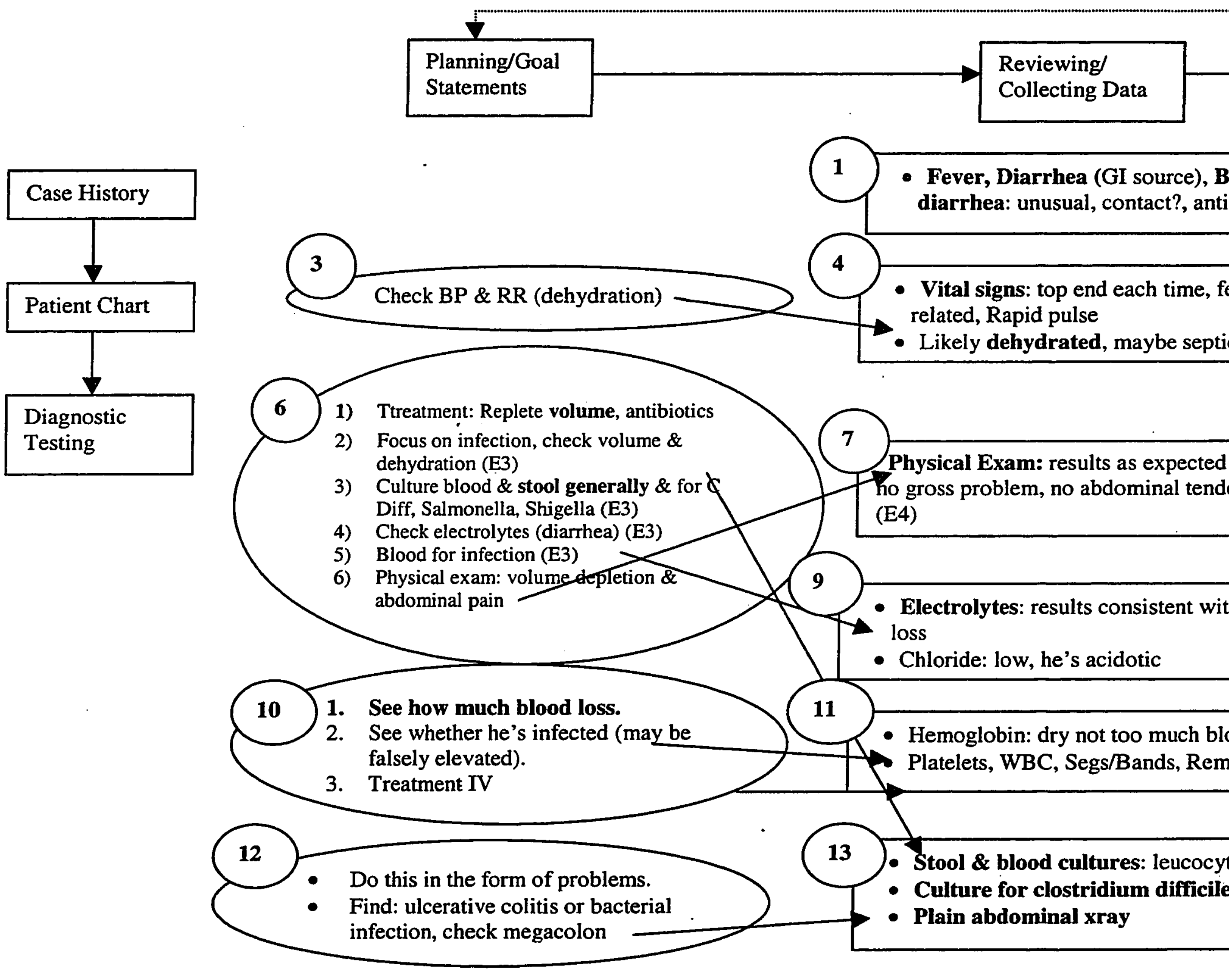
8
- Physical Exam: Anxiety, nothing to suggest Thyrotoxicosis
- **Small Bowel Follow Through (SBFT):**
  - no evidence of IBD,
  - evidence of malabsorption: normal B12, low folate

10
- **Antigliadin Antibodies:** not 100% sensitive
- Ferritin: very low just confirms iron deficiency

12
- **Gastroscopy**: malabsorption due to some other reason (Antigliadin is the gold standard)
- **Carotene**: low confirms malabsorption

13
Final Diagnosis: Some type of malabsorption syndrome (Sprue: unlikely due to negative results for Anti-Gliadin, Small Bowel Biopsy, & SBFT)

Figure 3. Expert Model for Case 2: Shigellosis



Reviewing/ Collecting Data

Hypotheses

1

- **Fever, Diarrhea** (GI source), **Bloody diarrhea**: unusual, contact?, antibiotics

2

1. **most common: a)** respiratory **infections**, 2) viral, 3) GI upsets
2. **(Some viruses), Salmonella, Shigella**
3. Clostridium Difficile (**colitis** & bloody diarrhea)
4. Amoebic, but no history, Inflammatory Bowel (young for it)
5. Infectious process MOST likely, bacterial
6. less common: medications, brain problems
7. worst possible: meningitis, septic shock

4

- **Vital signs**: top end each time, fever related, Rapid pulse
- Likely **dehydrated**, maybe septic shock

5

- Mechanical things like intussusception
- Aspirin can cause higher RR , could complicate main problem

7

Physical Exam: results as expected (E3), no gross problem, no abdominal tenderness (E4)

8

- Not Meningitis, ulcerative colitis unlikely

9

- **Electrolytes**: results consistent with volume loss
- Chloride: low, he's acidotic

11

- Hemoglobin: dry not too much blood loss
- Platelets, WBC, Segs/Bands, Remaining

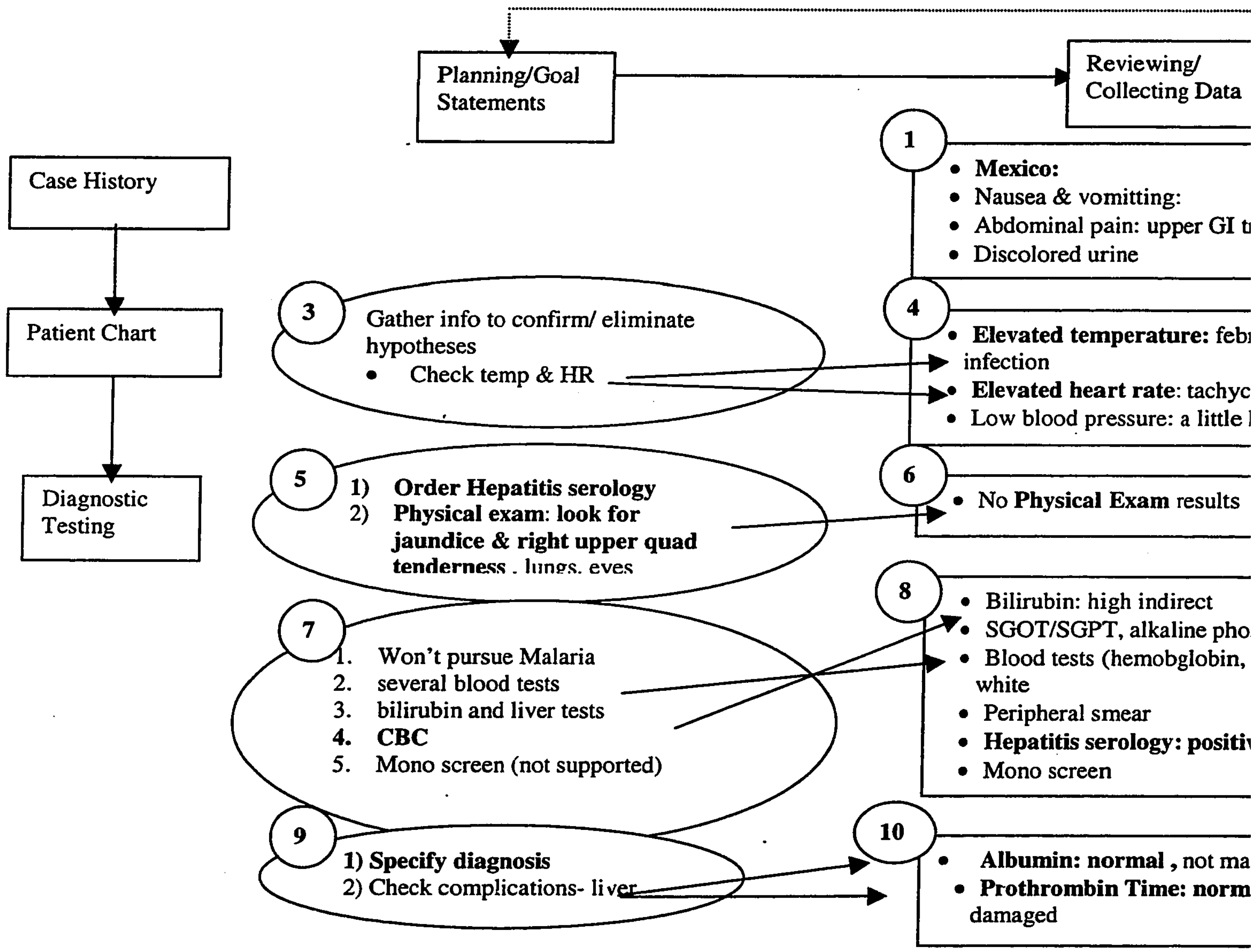
13

- **Stool & blood cultures**: leucocytes
- **Culture for clostridium difficile**
- **Plain abdominal xray**

14

Final Diagnosis: Bacterial Infection (Shigella, Salmonella, E-coli)

Figure 4. Expert Model for Case 3: Hepatitis A



Reviewing/
Collecting Data

Hypotheses

1

- **Mexico:**
- Nausea & vomitting:
- Abdominal pain: upper GI tract problem
- Discolored urine

2

- (Mexico) could be exotic disease or more common
- (nausea, vomiting) could be Tourista but no diarrhea
- Other infectious causes, malaria, hemolytic uremic syndrome
- Symptoms: more serious than Traveller's diarrhea

1. **Hepatitis A**, hepatitis B, Alcohol-induced hepatitis
2. **Drugs**
3. GI complaints
4. Gallstone (low on list)

4

- **Elevated temperature:** febrile, goes with infection
- **Elevated heart rate**: tachychardic, goes with temp
- Low blood pressure: a little hypotensive

6

- No **Physical Exam** results

8

- Bilirubin: high indirect
- SGOT/SGPT, alkaline phosphotase, LDH
- Blood tests (hemobglobin, platelets, WBC): high white
- Peripheral smear
- **Hepatitis serology: positive A,** negative B & C
- Mono screen

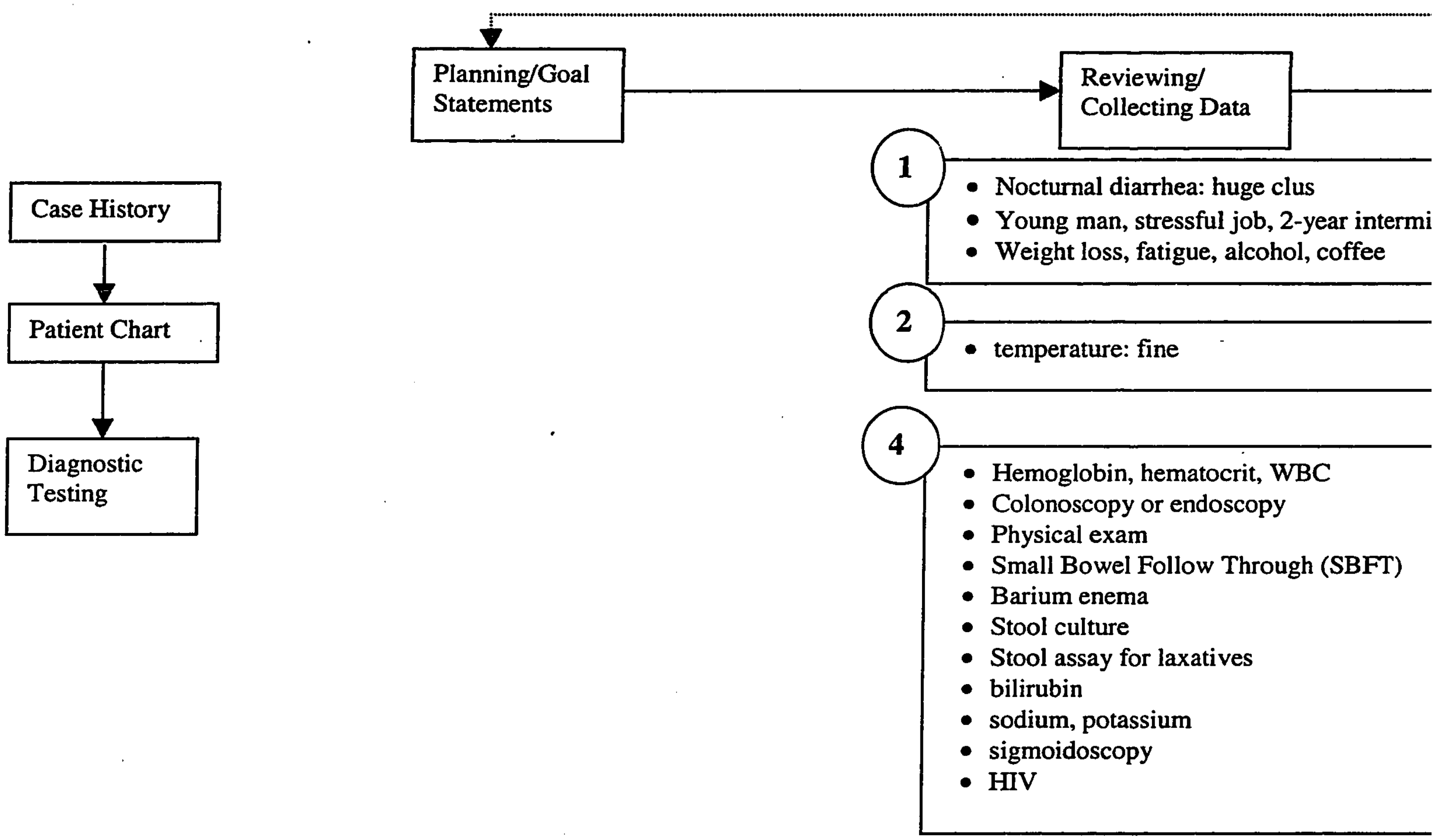
10

- **Albumin: normal ,** not massive necrosis
- **Prothrombin Time: normal,** liver not badly damaged

11

(Final Diagnosis): Hepatitis A

Figure 5. Sample Student Model: Participant S1 for Case 1: Celiacs



Reviewing/
Collecting Data

Hypotheses

- Nocturnal diarrhea: huge clus
- Young man, stressful job, 2-year intermittent diarrhea
- Weight loss, fatigue, alcohol, coffee

- temperature: fine

- Hemoglobin, hematocrit, WBC
- Colonoscopy or endoscopy
- Physical exam
- Small Bowel Follow Through (SBFT)
- Barium enema
- Stool culture
- Stool assay for laxatives
- bilirubin
- sodium, potassium
- sigmoidoscopy
- HIV

3

Crohn's or ulcerative colitis
(which one gives nocturnal
diarrhea?)

Coding

The coding scheme identifies a breakdown of operators and knowledge states (data and inferred) as described in the previous chapter. The way these operators and states were used by the two groups is discussed in detail. Several of the more interesting findings are highlighted in this section and linked to relevant research. The structure of the coding scheme and its implementation in N-Vivo allow for flexibility in examining the operators and knowledge states used by participants.

1. BioWorld Activity

The number of utterances per activity provides an indication of how much emphasis was placed on each activity by the groups. Table 10 provides a breakdown of the number of utterances by activity, case, and group. Students and residents were generally similar in their percentage of utterances by BioWorld case.

Table 10.

Number of Utterances by Activity, Case, and Group

Phase	Problem Solving			Post-Problem Reflection		
BioWorld Activity	1. Case History	2. Patient Chart	3. Diagnostic Testing	4. Categorizing	5. Prioritizing	6. Expert Trace
Students (n=10)						
Case 1	183	155	987	122	120	81
Case 2	135	189	855	139	146	27
Case 3	134	105	688	97	96	17
Sum	452	449	2530	358	362	125
M (overall)	15.07	14.97	84.33	11.93	12.07	4.17
SD	11.66	11.56	31.74	10.37	10.17	5.25
%	10.57	10.5	59.17	8.37	8.48	2.93
Residents (n=10)						
Case 1	152	126	1030	22	72	74
Case 2	117	145	874	8	105	27
Case 3	111	118	614	19	80	25
Sum	380	389	2518	49	257	126
M (overall)	12.67	12.97	83.93	1.63	8.57	4.2
SD	8.36	9.85	28.05	3.69	6.41	4.14
%	10.22	10.46	67.7	1.32	6.91	3.39
Experts (n=2)						
Case 1	203	28	319	0	20	31
Case 2	255	54	267	0	76	10
Case 3	136	21	105	0	13	0
Sum	594	103	691	0	109	41
M (overall)	99	17.17	115.2	0	18.17	6.83
SD	50.18	9.09	62.08	0	18.96	7.17
%	38.62	7.7	44.93	0	7.09	2.67

Problem solving versus post-problem reflection. All groups made many more utterances during the *Problem Solving* phase (students = 80.24%, residents = 83.38%, and experts = 91.25%) than they did during the *Post-Problem Reflection* phase. Working through the cases necessarily involves substantial processing to reach a solution. In contrast, the *Post-Problem Reflection* activities were obviously not critical to solving a BioWorld problem, and they were also unfamiliar tasks for the participants. As such, it is not surprising that the amount of activity was so much higher in problem solving as opposed to post-problem reflection.

Individual activities. All groups made more utterances during the Diagnostic Testing activity than any other (59.17% of utterances for students, 67.70% for residents, and 44.92% for experts). With the exception of #4 Categorizing, the students and residents had a similar overall % breakdown in the number of activities. More students than residents actually engaged in the #4 Categorizing task, and this accounts for their higher % of utterances for this activity (8.37 % as compared to 1.32%). Neither expert engaged in the Categorizing task.

Case history activity. The experts made nearly four times as many utterances overall during the Case History activity than either of the other groups. They made 38.62% of overall utterances during this activity as compared to students and residents (10.57% and 10.46%, respectively). As the expert models have indicated, the experts work on reviewing case information, generating hypotheses to account for many possibilities, and develop action plans early in the diagnostic process. They apparently engage in more and different processing to initially set up a problem. Many studies have reported that experts spend more time than sub-experts in analyzing problems "qualitatively" before attempting to solve them (Ericsson, 2003). The current results support this finding. For every case, the experts spent considerably longer than either of the other groups working on case history information before proceeding to solve the case. Subsequent analyses will shed more light on the processing that experts engaged in during the Case History activity.

Physical exam. Students and residents made roughly 10% of their utterances overall (similar to the case history) during this activity as compared to roughly 7.7% for experts. When the physical exam is reviewed, the experts have already reviewed the Case

History and worked on it extensively (as compared to the other groups). This may explain why they do not elaborate to the same extent on the Physical Exam data; instead, they may fit this new data into their elaborated representation of the case. If any surprising or inconsistent data were introduced in the physical exam, a different pattern of results may have emerged. On the other hand, students and residents made as many utterances in response to the Physical Exam data as they did in response to the Case History data. It seems that these groups may have processed all new information to a similar degree, building less elaborate representations of the cases. They also tended to give fewer diagnostic hypotheses (and less complete sets) in response to initial data, and may have offered more than experts in response to physical examination data.

Various studies have found different results in terms of the amount of time and/or processing participants required to complete various case components. In the study by Rikers et al. (2003), the researchers found that sub-experts and experts solving cardiology cases showed no proportional differences in the amount of time they spent working on different case components. All groups spent most of their time (roughly 2/3) on the case history and physical examination. The researchers suggest that less-experienced participants may have needed the time to process the information. Experts may instead have developed a more complex representation of the case. The current results would support these findings.

Diagnostic testing. It is not surprising that all groups generated the highest percentage of utterances for activity while working on diagnostic testing (e.g., ordering tests and obtaining results). Unlike the case history and physical exam, diagnostic testing in BioWorld is an interactive activity; participants are required to request specific test results. This partly explains why diagnostic testing generated more utterances by both groups and experts. However, as compared to experts, this activity generated a much higher percentage of utterances by students and residents. This may simply be explained by the experimental task, or may also indicate that this information is more important or useful to these groups. Subsequent analyses to indicate that diagnostic testing information versus case history information may be an important distinction as far as the data participants used to solve the cases. However, researchers including Rikers et al. (2003) have found that sub-experts and experts regard the history and physical exam as more important than the laboratory data or additional findings. If these components were also interactive in BioWorld, a different pattern of results may have been observed. Further, the experimental task may explain why more processing was devoted to the diagnostic testing activity in the current study and not in the Rikers et al. study. In the latter study, all data were presented directly to the participants. In the current study, the diagnostic testing data had to be collected by participants.

In sum, these analyses along with the expert models, suggest that while students and residents devoted roughly the same percentage of utterances to the case history and physical examination, experts engage in more elaborate initial processing including reviewing data, specifying hypotheses, and developing action plans.

2. Operator Use

The coded data were examined for the frequency of use for different operators and knowledge states. Frequency counts were calculated by case, group, and participant. The counts were tabulated using document searches in QSR N-Vivo. In order to minimize the effects of the length of different protocols, the percent frequency is used in reporting.

Table 11 provides a breakdown of operator used by participant group. It includes both the frequency and percentage of operator use by group (excluding the use of the operator "State"). Overall, students and residents were fairly similar in their percentage use of these operators, with larger differences in the use of the *Evaluate* and *Link* operators. The experts showed a somewhat different breakdown, with more use of *Evaluate* and *Link* operators (close to 70% of total) and less use of the *Request* operator as compared to students and residents. Operator use is briefly described here, and will be further examined in following sections.

2 1 State. State was used when a participant simply made a statement, such as a hypothesis. This operator was not coded in N-Vivo, although it was used extensively by participants. The reason is that this operator is essentially the "default".

2 2 Request. Requests made up the highest percentage of total operator use for both residents and students, but not for experts. For residents, requests accounted for more than half of all operator use (50.26%), while it accounted for somewhat less of total student operator use (44.72%). Further, both groups had a considerably higher percentage frequency use of this operator than did the experts (12.17%).

Table 11.

Breakdown of Operator Use (Frequency & Percentage) by Group

	Students (n=10)		Residents (n=10)		Experts (n=2)	
Operator	Freq	%	Freq	%	Freq	%
2 1 State*						
2 2 Request	585	44.72	686	50.26	65	12.17
2 3 Explain	53	4.05	29	2.12	50	9.36
2 4 Evaluate	369	28.21	276	20.22	175	32.77
2 5 Link	247	18.88	315	23.07	195	36.52
2 6 Read	54	4.13	59	4.32	49	9.18
Total	1308		1365		534	

The information requested by the groups was examined using Boolean searches in N-Vivo. Table 12 provides a breakdown of requests (frequency and percentage) by data type and participant group. Student, resident, and expert requests predominantly dealt with test results. Experts had a somewhat lower percentage (90%) as compared to both students (95.93%) and residents (97.52%). Experts also had a higher percentage of requests for case information (5.71%) than did the students (0.78%) or the residents (1.17%). Students made more requests for missing information (3.29%) than did the residents (1.31%). Expert requests for missing information were similar to those of students (2.86%). Finally, none of the groups requested library information.

Table 12.

Breakdown of Requests (Frequency & Percentage) by Data Type and Group

	Students (n=10)		Residents (n=10)		Experts (n=2)	
Request	Freq	%	Freq	%	Freq	%
3 1 1 Case Information	5	0.78	8	1.17	4	5.71
3 1 2 Test Result	613	95.93	669	97.52	63	90
3 1 3 Test	0	0	0	0	1	1.43
3 1 4 Missing	21	3.29	9	1.31	2	2.86
3 1 5 Library	0	0	0	0	0	0
Total	639		686		70	

This difference suggests that, while students and residents devote a larger portion of operator use to requests (approximately 45-50% excluding the *State* operator), experts engage more in other cognitive actions (e.g., explaining, linking, and evaluating). They also engaged in different types of processing prior to diagnostic testing as shown by the expert models. In contrast, the residents' and students' greater use of *Request* again partly reflects the nature of the experimental task (requests were required to obtain diagnostic testing data) and may also reflect a greater reliance on diagnostic testing data. Residents did use more diagnostic testing data during post-problem reflection activities than students or experts.

2 3 Explain. Students used this operator approximately twice as often as residents. However, for both groups, this still accounted for a fairly small percentage of overall operator use (4.04% and 2.12%, respectively). Experts use of this operator accounted for 9.36% of overall operator use. An N-Vivo search was conducted to determine which knowledge states were associated with the use of the *Explain* operator, and for all groups, most explanations related to *Strategic* states. The pattern of use was similar for students and residents. Students' higher percentage use of this operator included more frequent explanations of certain *Data* states (case history, test results), *Strategic* states (plans, actions), and one *Self* state (performance) as compared to residents. In contrast, 75% of experts' explanations were associated with *Strategic* states (plans, actions) with a few other explanations relating to *Data* states and *Self* states.

2 4 Evaluate. Table 11 also shows that students used the Evaluate operator more than residents (28.21% compared to 20.22%). Both groups used this operator less than did the experts (32.77%). Table 13 provides a further breakdown of the knowledge states that were used in conjunction with the Evaluate operator by participant groups. For all groups, a high percentage of requests dealt with diagnostic testing, which makes sense given the experimental task. However, some participants did request additional case information (none was provided). As the data indicate, the experts requested this information more frequently than students or residents.

2 5 Link. As reported in Table 11, students used this operator in a smaller percentage of utterances than residents (18.88% compared to 23.07% of total operator use). The experts used this operator more than any other operator, and considerably more (36.52%) than either of the other two groups. The larger proportion of linking actions by experts is consistent with their more elaborate processing and the relationship between data, hypotheses, and plans that was identified in the expert models. For example, in several instances the initial review of the case history information by experts led them to link certain data items to hypotheses and to plans for diagnostic testing. These plans were designed to enable them to distinguish between competing hypotheses. In sum, the greater use of this operator by experts may suggest that they put more emphasis than the other groups on developing a coherent understanding of the data as a whole.

Table 13.

Breakdown of the Use of Evaluation (Frequency & Percentage) by Data Type and Group

	Students (n=10)		Residents (n=10)		Experts (n=2)	
Knowledge States	Freq	%	Freq	%	Freq	%
3 1 Data States						
3 1 1 Case Information	231	59.69	189	66.55	98	56.65
3 1 2 Test Result	55	14.21	30	10.56	31	17.92
3 1 3 Test	11	2.84	4	1.41	9	5.20
3 1 4 Missing	0	0	1	0.35	2	1.16
3 1 5 Library	3	.78	0	0	0	0
3 2 Inferred States						
3 2 1 Patient Condition	7	1.81	8	2.82	4	2.31
3 2 2 Diagnostic Solutions						
3 2 2 1 Hyp-new	17	4.39	5	1.76	4	2.31
3 2 2 2 Hyp	24	6.20	16	5.63	12	6.93
3 2 2 3 Final Diagnosis	10	2.58	2	.70	0	0
3 2 3 Strategic						
3 2 3 1 Plan*	1	0.26	3	1.06	3	1.73
3 2 3 1 1 Action	1	0.26	0	0	1	0.58
3 2 3 1 2 Potential Action	0	0	0	0	0	0
3 2 3 1 3 Treatment	0	0	0	0	0	0
3 2 4 Self						
3 2 4 1 Knowledge	20	5.17	14	4.93	8	4.62
3 2 4 2 Performance	7	1.81	12	4.23	1	0.58
3 2 4 3 Regulation	0	0	0	0	0	0
3 2 5 Relevant	0	0	0	0	0	0
Total	387		284		173	

Table 14 provides a further breakdown of the knowledge states that were used in conjunction with the *Link* operator by participant groups. The table lists the frequency and percentage frequency of the use of each knowledge state occurring with the *Link* operator. The use of *Link* with the *Test Result* state is one where a difference between the groups exists (students = 20.2%, residents = 14.73%). Other relatively small differences between the groups are also shown in Table 14. For the experts, the breakdown of the use of this operator is similar to the other groups, with the exception of 1) lower frequency percentages for *Case Information* (25.19%), *Hyp-New* (10.33%), and *Final Diagnosis* (0.5%), and 2) a higher frequency percentage for *Hyp* (23.43%).

Table 14.

Breakdown of the Use of Link (Frequency & Percentage) by Data Type and Group

	Students (n=10)		Residents (n=10)		Experts (n=2)	
Knowledge States	Freq	%	Freq	%	Freq	%
3 1 Data States						
3 1 1 Case Information	151	29.9	191	29.94	100	25.19
3 1 2 Test Result	102	20.2	94	14.73	70	17.63
3 1 3 Test	5	0.99	17	2.66	5	1.26
3 1 4 Missing	1	0.2	2	0.31	3	0.76
3 1 5 Library	1	0.2	0	0	0	0
3 2 Inferred States						
3 2 1 Patient Condition	77	15.25	93	14.58	59	14.86
3 2 2 Diagnostic Solutions						
3 2 2 1 Hyp-new	67	13.27	90	14.11	41	10.33
3 2 2 2 Hyp	82	16.24	103	16.14	93	23.43
3 2 2 3 Final Diagnosis	8	1.58	20	3.13	2	0.5
3 2 3 Strategic						
3 2 3 1 Plan*	3	0.59	8	1.25	12	3.02
3 2 3 1 1 Action	5	0.99	12	1.88	3	0.76
3 2 3 1 2 Potential Action	3	0.59	5	0.78	4	1.01
3 2 3 1 3 Treatment	0	0	2	0.31	5	1.26
3 2 4 Self						
3 2 4 1 Knowledge	0	0	1	0.16	0	0
3 2 4 2 Performance	0	0	0	0	0	0
3 2 4 3 Regulation	0	0	0	0	0	0
3 2 5 Relevant	0	0	0	0	0	0
Total	505		638		397	

Further, students and residents used the *Link* operator slightly more with previously mentioned hypotheses (coded *Hyp*, ~16%) than with new hypotheses (*Hyp-new*~14%). For the experts, previously mentioned hypotheses (*Hyp*) were linked much more often (~23%) and new hypotheses were linked less often (~ 10%). This finding might be explained by referring back to how the experts proceeded with the case information. As they expert models show, they tended to generate a range of hypotheses early, in response to case history information. Some early hypotheses were linked to other knowledge states, but more often the linking of diagnostic hypotheses occurred with previously mentioned hypotheses. Other studies also suggest that linking is characteristic of experts. For example, Joseph and Patel (1990) found that high-domain knowledge (HDK) participants generated links to organize textual case descriptions, while less-experienced personnel did not.

2 6 *Read.* Students and residents used this operator with a very similar frequency (4.12 to 4.32%, respectively). Experts used it more frequently (9.18%), which is explained by their periodic re-reading of case information.

Other Studies and Operator Use

In the case of operators, residents on average used *Request* and *Link* more and *Explain* and *Evaluate* less than students did. In contrast, experts made more use of the *Link* and *Evaluate* operators.

The experts' use of linking is consistent with results from other studies. Several researchers have established that expert knowledge is richer, more elaborated, and more interconnected than that of non-experts. For example, Joseph and Patel (1990) found that experts generated more links to relate relevant cues. Lesgold et al. (1988) also found that experts generated a greater number of findings and related them to other findings. In addition, in the two studies by Rikers et al. (2002a, 2002b), the researchers found that that expert explanations of symptoms and signs matched more concepts and links contained in canonical models than those generated by less experienced participants. Finally, Patel, et al. (1997) determined that experts formed richly integrated knowledge structures based on the history taking and used them throughout problem solving. Intermediates had less-integrated knowledge, while students' knowledge only superficially resembled that of experts. While the exact meaning of linking differs in these studies, the essential idea that expertise involves making or using more links between key concepts seems supported by these studies and the current study as well.

Norman et al. (1994) and others have also reported that the knowledge of experts is more richly elaborated than that of non-experts. Van de Wiel et al. (1995) also reported that expert explanations of medical concepts were more elaborate and fluent than those of non-experts. Investigating disease categories and prototypicality, Bordage and Zacks (1984) noted richer and more tightly networked knowledge in experts that was made possible by a greater number of connections between the various disorders within categories. Finally, Bordage and Lemieux (1991) found that experts had deeper problem representations and used more elaborated abstract knowledge to solve problems. These results would appear to be supported by the findings of this study.

3. Knowledge State Use

Table 15 presents the frequency and percentage of the use of each state by group.

Table 15.

Breakdown of Knowledge State Use (Frequency and Percentage) by Group

Knowledge	Students (n=10)		Residents (n=10)		Experts (n=2)	
State	Freq	%	Freq	%	Freq	%
3 1 Data States						
3 1 1 Case Information	704	26.08	663	24.49	287	23.84
3 1 2 Test Result	1038	38.46	1064	39.31	207	17.19
3 1 3 Test	37	1.37	56	2.07	23	1.91
3 1 4 Missing	54	2	53	1.96	27	2.24
3 1 5 Library	9	0.33	0	0	0	0
3 2 Inferred States						
3 2 1 Patient Condition	128	4.74	154	5.69	97	8.06
3 2 2 Diagnostic Solutions						
3 2 2 1 Hyp-new	168	6.22	183	6.76	95	7.89
3 2 2 2 Hyp	187	6.93	173	6.39	144	11.96
3 2 2 3 Final Diagnosis	48	1.78	55	2.03	7	0.58
3 2 3 Strategic						
3 2 3 1 Plan*	15	0.56	20	0.74	43	3.57
3 2 3 1 1 Action	136	5.04	136	5.02	149	12.38
3 2 3 1 2 Potential Action	76	2.82	46	1.7	42	3.49
3 2 3 1 3 Treatment	5	0.19	22	0.81	18	1.5
3 2 4 Self						
3 2 4 1 Knowledge	29	1.07	29	1.07	13	1.08
3 2 4 2 Performance	16	0.59	20	0.74	2	0.17
3 2 4 3 Regulation	27	1	17	0.63	3	0.25
3 2 5 Relevant	22	0.82	16	0.59	47	3.9
Total	2699		2707		1204	
Adjusted (-15, -20, -43)	2684		2687		1161	

Knowledge States: Data

These states comprise the raw data, such as case history information (e.g., symptoms), physical examination information (e.g., vital signs), and diagnostic test results.

3 1 1 Case Information. Approximately 25% of knowledge state coding was for Case Information, for all of the groups.

3 1 2 Test Result. Students and residents used this type of knowledge state to a similar extent (38.46% and 39.31% respectively). In contrast, experts used test results to a lesser extent (17.19%). Again, compared to the students and residents, experts processed a lot of information prior to testing.

3 1 3 Test. The three groups similarly used this knowledge state to a very limited degree (approximately 1-2%).

3 1 4 Missing. Missing or unavailable knowledge states were mentioned in approximately 2% of knowledge states for each group.

3 1 5 Library. Only students used knowledge states relating to the BioWorld library. The frequency (9) and percentage are very low (0.33%). The participants rarely accessed library information.

Knowledge States: Inferred

These states consist of information that has been gained through some form of inference. That is, the participant has performed some cognitive operation to modify or interpret raw data.

3 2 1 Patient Condition. The students and residents referred to a patient's physiological condition with similar frequency (4.74% and 5.69%, respectively). Experts used *Patient Condition* somewhat more frequently (8.06%).

3 2 2 Diagnostic Solutions. This category includes potential solutions (new hypotheses and previously mentioned hypotheses) and final diagnoses.

3 2 2 1 Hyp-New: Hyp-new refers to a diagnostic hypothesis not mentioned previously. For students and residents, this knowledge state was used to a similar extent (6.22% and 6.76%), while experts used them to a slightly higher extent (7.89%). This translates to an average of 16.8 new hypotheses generated by students and 18.3 by residents over all of the cases. Experts generated an average of 40.5 new hypotheses over all of the cases. The generation of new hypotheses was also examined in terms of where in the BioWorld cases the new hypotheses were generated. As shown in Table 16, students generated the largest number of new hypotheses in response to Case History information, while the residents generated more during the Diagnostic Testing activity. Not surprisingly, the experts generated a large majority of their new hypotheses in response to Case History information. They generated very few new hypotheses in response to Patient Chart information.

Table 16.

Number of New Hypotheses Generated by Activity and Group

	Case History	Patient Chart	Diagnostic Testing
Students (n = 10)	68 (40.5%)	48 (28.6%)	51 (30.4%)
Residents (n = 10)	53 (29%)	51 (27.9%)	77 (42.1%)
Experts (n = 2)	81 (85%)	2 (2%)	11 (11.6%)

Frequently, students and residents discussed the Case History information and Patient Chart results together, after viewing the Patient Chart. Some of their hypotheses, although generated during or after the Patient Chart activity, actually deal with case history information. So, it may be more meaningful to collapse the two activities. In that case, there is still a discrepancy between these groups and the experts, who generate roughly 87% of new hypotheses in response to this information and the remainder in response to testing. The other two groups generated a smaller percentage of new hypotheses in response to this information and substantially more new hypotheses in response to Testing data. In four instances, new hypotheses were generated during post-problem reflection.

A two-factor (group and activity) Analysis of Variance (ANOVA) with repeated measures on one factor (activity) determined that there were no significant differences in the number of hypotheses between the groups [$F(1,18) = 0.16$, $p = 0.69$]or across the activities [$F(2,36) = 0.52$, $p = 0.60$]. A relatively large amount of variability within the groups accounts for these findings. In fact, for both students during the Case History activity and for residents during Diagnostic Testing, a high number of new hypotheses were generated by a few individuals only. Based on these results, the students and residents did not differ overall in terms of the activities in which they generated new hypotheses. However, it is interesting to note that both of the experts generated a very large proportion of their new hypotheses in response to case history information. Finally, as other analyses show, the students more often rated case information as important to making a diagnosis, while residents more often reported testing information as important.

The results support findings from other studies that show that the generation of correct hypotheses early is a feature of expertise and diagnostic success (e.g., Gruppen et al., 1988, 1993; Johnson et al., 1981; Neufeld et al., 1981). The experts used the Case History information to generate many more hypotheses on average than either students or residents, and the experts generated 85% of their new diagnostic hypotheses in response to Case History information. The expert models also demonstrate that the Case History information leads to multiple hypotheses, and initial plans for determining whether the hypotheses are correct.

Further, Patel et al. (1984) found that after generating the correct diagnosis, experts generated few additional diagnoses and worked on ruling out and confirming different diagnoses, and identifying secondary problems. In contrast, non-experts continued to generate new diagnostic hypotheses, and were often not successful in diagnosing the cases. The current results generally fit these findings.

An N-Vivo search was conducted on *hyp-new* to determine at what point during problem solving the correct hypothesis was generated for each of the groups. For Case 1 (difficult), two students and two residents generated the correct diagnosis in response to Case History information. For Case 2 (moderately difficult), two students and one resident specifically named Shigellosis, while a number of them gave a correct general diagnosis of an acute infectious process. For Case 3 (easy) six students and five residents generated the correct diagnosis in response to case history information. In contrast, both experts generated the correct hypothesis in response to case history information for each case. They specifically named Celiac disease and Shigellosis in their final diagnoses. These findings would again point to the more elaborate processing of experts that resulted in early identification of the correct hypothesis for all cases. The results also suggest that, with decreasing case difficulty, the students and residents may also increase the number of correct hypotheses generated early in response to case history information.

The Gruppen et al. (1988) study determined that a high percentage of physicians generated the correct diagnosis (amongst others) in response to only the chief complaint and personal information. When case history information was included, the correct diagnosis was generated in over 90% of cases. The results of this study seem to lend further support to the finding that diagnostic success is usually accompanied by the early generation of the correct hypothesis.

Further, other studies have also demonstrated that experts typically generate the correct diagnosis early in the problem solving process (Neufeld et al., 1981). Other researchers have concluded that if the correct hypothesis is not generated (amongst others) early in the process, there is a very good chance the case will not be correctly diagnosed. In this study, for all cases, the experts generated the "correct" hypothesis, amongst others, after viewing this initial information.

Allen et al. (1998) investigated evidence gathering, and found that using adequate evidence was a function of the early development of accurate hypotheses. Briefly, experts generated accurate hypotheses early, and carried out a controlled process of refining initial hypotheses to resolve any inconsistent data. In this study, the lengthy protocols of experts during the case history activity apparently reflects some of this early review of the evidence and refinement of diagnostic hypotheses.

3 2 2 2 Hyp: Hyp refers to a diagnostic hypothesis mentioned previously. Students and residents used *Hyp* to a similar extent (6.93% to 6.39%), while experts used them in a higher percentage of utterances (11.96%). This may be more evidence of elaborate processing on the part of experts, as they apparently mentioned many potential hypotheses early, and later mentioned them frequently.

3 2 2 3 Final Diagnosis: This code was applied to each final solution for a case. Again, students (1.78%) and residents (2.03%) showed a similar usage, while expert usage was somewhat lower (0.58%). The accuracy of the final diagnoses was rated by an expert (as previously discussed). The final diagnoses were also examined in conjunction with what part of the problem solving activity they were generated in. For Case 1, only two residents correctly diagnosed the case. One of them was a resident who generated the correct diagnosis in response to Case History information. The other resident first generated the correct diagnosis during diagnostic testing. The one resident and two students who mentioned the correct diagnosis in response to Case History information all failed to diagnose the case.

3 2 3 Strategic. The plan and goal statements made by participants are examined I some detail in this section. They were found to be a feature of expert performance, appearing multiple times in each expert protocol, while they do not appear consistently in the protocols of the other groups. The analysis includes the number of Plans generated by individuals and groups, the content of the Plans, and the number of steps per plan. As defined previously, each plan consists of a series of goal statements (2 or more).

3 2 3 1 Plan. Multiple goal statements occurring together were coded as plans. The use of strategic states is examined in greater detail in the following section (*Use of Strategic Thinking*).

3 2 3 1 1 Action. Again, students (5.04%) and residents (5.02%) stated actions to a similar extent. Experts used them more frequently (12.38%).

3 2 3 1 2 Potential Action. These knowledge states were used to a similar, limited extent by each group (students = 2.82%, residents = 1.7%, and experts = 3.49%).

3 2 3 1 3 Treatment. Finally, treatment states were used very little by students and residents (0.19% and only slightly more by residents (0.81%), and experts (1.5%).

As mentioned previously, planning or goal setting has not been a main focus of medical expertise research. It has been addressed to some extent in other sub-domains of medicine (for an example, see Xiao et al., 1994). This is partly due, no doubt, to the fact that medical diagnosis is a problem solving activity that has an established general structure. That is, virtually every time a diagnosis is made, the diagnostician proceeds through a series of basic steps (Case History, Physical Exam, etc.). At the outset of problem solving, diagnosticians approach the task with a general goal of determining the cause of patient's complaint using appropriate methods (e.g, diagnostic testing). The actual solution process then proceeds as the specific features of the case are worked through.

In this context, strategic thinking then amounts to working within a general framework: specifying localized goals, taking steps to achieve those goals, and specifying further goals. Therefore, in the case of BioWorld, the results show that expert strategic thinking is localized and prevalent during problem solving, and responsive to incoming data. The data also show that the experts engage in more extensive examination of the initial data (Case History information) prior to proceeding with the case, perhaps generating a more detailed problem representation.

As defined in the coding scheme, a plan consists of multiple intended or potential actions or goals. Table 17 presents the mean number of plans generated by individuals and groups. Experts clearly generated more plans than the other participants, and they generated multiple plans for each case. In contrast, most students (8/10) and residents (6/10) generated either no plans or 1 plan, while the rest generated multiple plans. Finally, the number of plans ranged from 0 to 5 for students and between 0 and 3 for residents. With the exception of residents solving Case 2, there is a linear decrease in the number of plans generated with decreasing case difficulty. Interestingly, even for the easy case, both experts still generated multiple plans, while only three other participants generated one multi-step plan each.

Not surprisingly, most plans were generated during the diagnostic testing phase. However, it is interesting to note that experts also generated 8 plans in response to case history information. Five out of six expert protocols contained at least one plan based on case history information. Residents generated 1 plan and students generated 3 plans based on the case history.

The number of steps in each plan was also tabulated and the results are shown in Table 17. The table represents each participant, including the number of plans and the number of steps per plan that s/he generated for each case. Overall, the groups had similar mean number of steps per plan. Most student plans (12/15) ranged from 2-4 steps, residents' plans mainly consisted of 2-3 steps (16/20), and almost all (42/43) of the experts' plans contained 2-6 steps. Both experts generated multiple plans for each case, while students and residents did so for only a small number of cases (4/30 or 13.33% and 7/30 or 23.33%, respectively). Finally, the data suggest that case difficulty may be a factor in plan generation.

Most students and residents developed plans (7 students, and 8 residents). The number of plans per participant differed for the two groups. For students, the majority of plans (10/15) were generated by two participants only (S5 and S8), each of whom generated five plans. Five other students generated one plan each. In contrast, six residents generated two or more plans, while two of them generated one plan each. Clearly, there is no "optimal" number of plans for problem solving in this context. This analysis is useful to show the variety of responses from individuals and groups. Further, it demonstrates that the mean number of steps per plan is relatively low.

Table 17.

Number of Plans and Steps Per Plan Generated by Individual, Group, and Case

	Case 1: Celiac (difficult)		Case 2: Shigellosis (medium)		Case 3: Hepatitis A (easy)		# Plans, # Steps
	# plans	#steps	# plans	# steps	# plans	# steps	
Students							
S1	-	-	-	-	-	-	
S2	1	4	-	-	-	-	
S3	1	2	-	-	-	-	
S4	-	-	1	2	-	-	
S5	2	2, 4	2	3,11	1	2	
S6	1	4	-	-	-	-	
S7	-	-	-	-	-	-	
S8	2	5, 6	2	2,3	1	3	
S9	-	-	-	-	-	-	
S10	-	-	1	2	-	-	
Total	7	27	6	23	2	5	15
Mean #	0.7	3.86	0.6	3.83	0.2	2.5	3.67
SD	0.82		0.84		0.42		
Residents							
R1	-	-	-	-	-	-	
R2	-	-	-	-	-	-	
R3	1	4	-	-	-	-	
R4	-	-	1	2	-	-	
R5	2	6,2	-	-	-	-	
R6	2	5,2	2	3,3	-	-	
R7	-	-	2	2,3	-	-	
R8	3	3,2,5	2	3,3	1	3	
R9	-	-	2	2,3	-	-	
R10	1	2	1	2	-	-	
Totals	9	31	10	26	1	3	20
Mean #	0.9	3.44	1.0	2.6	.10	3	3.0
SD	1.10		0.94		0.95		
Experts							
E3	9	2,2,3,6,4, 2,2,5,5	5	3,3,9,2,4	5	2,3,4,4,3	
E4	8	2,3,4,2,2, 2,6,2	11	4,4,4,2,5,3, 6,3,4,4,3	5	6,2,2,3,6	
Totals	17	54	16	63	10	35	43
Mean #	8.5	3.17	8	3.94	5	3.5	3.53*
SD	0.71		4.24		0		

The content of plans was examined in order to determine what differences might exist in what the groups generated plans about (testing, treatment, gathering more information, a diagnostic list, and general strategies). Table 18 provides a breakdown of the content of plans by participant, group, and case.

Table 18.

Breakdown of Plans Based on Content

	Testing	Treatment	Gather Info	List	General
Students					
S1	-	-	-	-	-
S2	1	-	-	-	-
S3	1	-	-	-	-
S4	1	-	-	-	-
S5	2	1		1	1
S6	-	-	1	-	-
S7	-	-		-	-
S8	-	-	5	-	-
S9	-	-	-	-	-
S10	-	-	-	1	-
Total	5	1	6	2	1
Residents					
R1	-	-	-	-	-
R2	-	-	-	-	-
R3	-	1	-	-	-
R4	1	-	-	-	-
R5	1	-	-	1	-
R6	3	1	-	-	
R7	1	1	-	-	-
R8	3	3	-	-	-
R9	1	-	1	-	-
R10	2	-	-	-	-
Total	12	6	1	1	0
Experts					
E3	12	-	2	1	2
E4	13	5	2	-	1
Total	25	5	4	1	3

Most residents (7/10) generated plans related to Testing, and several more developed plans related to Treatment (4/10). Only two other resident plans were generated. Finally, the dispersion of plans across the categories is similar for the two experts, with the exception of the Treatment category.

For students and residents, more participants generated plans about testing than other types of content. However, one student did generate 5 plans to gather further information, which makes the student total number of plans higher for that category than for testing. Residents generated more testing plans (12) than did students (5). In terms of treatment, residents made more plans (6) than students (1), perhaps reflecting the residents' greater experience with patients. That is, they are more experienced with real patients, and this may be why they spontaneously provided treatment plans when not specifically asked to do so. One expert also provided 5 plans relating to treatment.

The analyses also revealed differences between students and residents in the use of strategic knowledge states. Students developed fewer plans than residents and the content of their plans was somewhat different. Student plans were divided among testing and gathering more information, while resident plans mainly concerned testing and, to a lesser extent, treatment. Not all students or residents generated plans, while experts generated multiple plans for each case. Experts generated many more plans, primarily concerning testing. Finally, the number of plans generated may be influenced by case difficulty or other case features; students, residents, and experts generated the fewest plans in response to Case 3 (easy). However, unlike the other groups, the experts did generate multiple plans for Case 3.

As explained in Chapter 2, studies of medical expertise have typically not focused specifically on planning as an important aspect of diagnosis. There are several possible explanations. First, medical diagnosis does not require *advanced* planning as some activities do (e.g., writing). Rather, the planning takes place in response to the particular features of a case. Second, medical diagnosis has a well-established process (conduct a history, then a physical examination, followed by diagnostic testing) that practitioners apply to reach a conclusion in virtually all instances of the activity. Third, the experimental tasks employed in expertise studies frequently have not elicited realistic problem solving including planning. Finally, as researchers in this area have noted, planning has, in general, not received much research attention. Expertise research including studies of diagnostic problem solving in medicine have focused on a wide range of other cognitive factors (processes, structures, knowledge types) as identified in Chapter 2. A few of the studies that have discussed planning in medical diagnosis contexts are discussed here.

Lesgold et al. (1988) found that experts are quick to start using general plans after viewing the case information (in their study, radiographs). After seeing the radiographs, experts quickly mentioned a diagnostic category and proceeded to discuss the case in that context (relating findings to the case). As the researchers found, experts in this study also started using the information as they encountered it. They generated more plans than other participants, and generated them in response to case history information in almost all cases. Importantly, experts generated plans concerning subsequent testing far more often than did the other groups, apparently using case information to guide their problem solving.

In the Lesgold et al. (1988) study, residents were less accurate in relating findings to diagnostic categories. Another characteristic of expertise these researchers identified is *opportunism*, or taking new possibilities into account as new information emerges. This includes both incorporating new data as it is obtained, and noticing new possibilities in existing data. In the current study, participants continued to receive new data as they progressed through the cases. As the models demonstrate, the experts apparently engaged in a cyclical process of interpreting, hypothesizing, and planning. In this way, their planning was based on data and seems to have proceeded in a systematic way. In contrast, the limited planning by the other participants suggests that they were likely not working in the same manner.

Patel, Groen, and Patel (1997) studied the performance of medical students, housestaff, and physicians during a "Patient Workup" task (history taking, interpretation of physical exam results, ordering and interpreting tests, providing management plans, and explanation of underlying pathophysiology). They found that participants used two different strategies for ordering tests: 1) a *structured* approach which involved relating tests to differential diagnoses, and 2) an *unstructured* approach whereby either no reason was given for ordering specific tests or there was no relationship between tests and potential diagnoses. They found that the structured approach was characteristic of experts. In the current study, experts did consistently provide plans for testing, which were developed in response to either or both case information and hypotheses. Planning and related knowledge states will be discussed further in subsequent sections.

A number of studies that relate to the use of strategic thinking in medical diagnosis have focused on the distinction between hypothesis-driven and data-driven problem solving. Typically, they reported that experts use data-driven problem solving, while less experienced individuals often employ hypothesis-driven or mixed strategies. The expert models demonstrate how the experts used the data in conjunction with hypotheses, plans, and testing. Their purposeful, cyclical use of case information, hypotheses, and planning apparently differs from the problem solving of less experienced groups.

3 2 4 Self. This category includes utterances that participants made concerning themselves, including their knowledge, performance, and self-regulation.

3 2 4 1 Knowledge. The three groups were almost identical in the percentage of use of this state type (approximately 1% for all). Students and residents both generated 29 utterances concerning their own knowledge. Almost all of these utterances by students (28) and residents (26) were statements indicating that the participant either did not know or could not remember certain information (e.g., "but then I can't remember what are all the symptoms with that" (R15- Case 2)). Experts generated 13 utterances concerning their own knowledge, and 11 of them concerned the fact that Case 2 involved a pediatric case. For all groups, the remaining utterances were statements concerning something the participant did know.

3 2 4 2 Performance. The frequency of use was very low for this state type (<1%) for each group. This code was applied 38 times in 25 protocols (1 Expert = 2, 7 Students = 16, 6 Residents = 20). For students and residents, half of these utterances concerning their performance were explanations [e.g., "These are basically my differential (diagnoses) I'm going through." (S5-Case 2)]. Participants in each group also gave several evaluations of their performance as positive or negative [e.g., "I guess I wasn't too bad"(S10-Case 1)], and identified several things they had forgotten to do [e.g., "Oh I should have done a urinalysis" (S1-Case 3)].

3 2 4 3 Regulation. The frequency was also very low for this state type (=/<1%) for each group. Students made more self-regulation utterances (27) than the residents (17). These utterances were of two basic types: utterances roughly corresponding to self-questioning about how to proceed (e.g., "What should I do next?") and utterances concerning the importance of certain case information (e.g., "What is important here?). Students made nearly three times as many utterances (20) concerning what to do next as they did about what was important (7). Residents made fewer (10) utterances concerning what to do next, and the same number (7) of utterances concerning what was important. Experts made regulatory utterances (3) about what was important only.

The use of the *Self-Regulation* knowledge states is potentially interesting as both students and residents spontaneously generated statements and questions about what to do next, and what information was important. Self-regulation is being investigated by many cognitive and medical researchers who are interested in improving learning (see Pintrich, 2000; Winne, 1997). Developing self-directed learning capacities is in fact a general goal for PBL-based medical training, as Evenson, Salisbury-Glennon, and Glenn (2001) have noted. These researchers have developed a situated model of self-regulation in medical school. They introduce the notion of *stance,* which refers to the form of self-regulatory actions that a learner adopts. They describe five stances they observed in medical students, which reflect a range of personal and environmental factors. These stances could be important in determining what is learned by students as they take part in problem-based learning activities such as solving BioWorld cases.

3 2 5 Relevant. Students and residents used these states to a very limited extent (22 times or 0.82% and 16 times or 0.59%, respectively). For students, almost all of these utterances (20) were single utterances, as were many of the residents' (11). Students generated a few (2) multi-utterance discussions of relevant information while residents generated a few more (5).

Experts generated utterances concerning relevant information somewhat more frequently (47 times or 3.9%) and generated more (16) multi-utterance discussions of relevant information (e.g., "Clostridium difficile infects your colon after you've had antibiotics. It's associated with antiobiotic infection. Antibiotics suppress your normal colonic flora and allow clostridium difficile to grow", E4-Case 2).

Summary

These analyses include a number of interesting findings, several of which are summarized briefly here. The expert models were presented as a characterization of the experts' cognitive processing. The models depict a cyclical data-plan-test process, which apparently does not reliably occur in the protocols of students or residents. They also demonstrate that experts spent considerably more time and made more utterances working through the case history information than did the other groups.

The results of coding scheme application have provided additional insights into diagnostic problem solving. Coding for BioWorld activity determined that participants devoted much more processing to problem solving activities rather than post-problem reflection. In line with various other studies, diagnostic testing was the activity where the groups concentrated most of their efforts.

The use of various operators and knowledge states by the groups has been examined. In terms of operators, students' and residents' utterances included a large proportion of requests typically for diagnostic test results. To a lesser extent they used evaluating and linking operations. In contrast, experts produced a larger proportion of linking operations than any other type, as well as more evaluating and more explaining utterances than the other groups. Experts also produced a much smaller proportion of requests than did the other groups. In sum, experts engaged in lengthier and more varied types of processing and apparently developed more coherent representations.

In terms of knowledge states, experts generated a large number of new hypotheses (Hyp-new) early in the diagnostic process. Students and residents generated less, but the results suggest that with decreasing case difficulty, the students and residents may increase the number of correct hypotheses generated early in response to case history information. In terms of planning, students and residents were variable, sometimes developing plans, and apparently developing fewer plans for the easy case than for the more difficult ones. The experts consistently developed multiple plans per case, particularly plans for subsequent testing actions. The residents resembled the experts more than the students did in terms of the number and content of their plans.

Post-Problem Reflection

After the BioWorld user has "solved" or diagnosed a case, s/he is directed to complete these activities in the following order: Categorizing, Prioritizing, and Expert Trace. This section will examine how participants completed these activities. Data for this portion of the analyses were taken from the forms that the participants worked with during these tasks and verbal protocol data. Verbalizations were typically very scant, while participants wrote their answers and other pertinent information on the materials provided.

Answering Research Question #3

3. ***How does the post-problem reflection (categorizing data, prioritizing data, and comparing to an expert list) of students and residents differ? How do their reflective actions compare with those of experts?***

Categorizing

Several participants indicated that this activity was not particularly helpful for them. This task was not completed by many of the participants. Not surprisingly, several suggested that this task really would not help them to solve the cases or was not realistic for them. As such, it will not be discussed further at this point.

Prioritizing

Types of Information in the Lists: Case History and Testing Items

The prioritized lists produced by participants (in writing) were examined and the results are summarized in this section. Briefly, lists were examined to determine 1) what types of information were used, and 2) what differences might exist between the groups in the contents of the lists. The lists were organized such that the items were written in decreasing order of importance. Therefore, the most important item(s) on a list would be rated "1", and subsequent items would be numbered in increasing order (2, 3, etc.). Participants were free to include as many or as few items as they pleased. The data were first tabulated by the individual participant, and subsequently collapsed in to summarize the data for the groups. For each case, the use of case history versus testing information by students, residents, and experts is compared. It is important to recall that there were only two experts, while there were ten participants in each of the other groups.

In addition, it should be noted that participants sometimes grouped several "pieces" of information together and gave the group a single rating. For example, the first item on participant R8's list for Case 1 included two test results (low folic acid and blood count) and a symptom (fatigue). The list items also varied in terms of generality; some items were specific or detailed (e.g., a specific test result such as low hemoglobin) while others were more general (e.g., test results). In addition, a few participants chose to list items using their general type (e.g., "Case History") while others listed test results using their type (e.g., "Radiology", "cultures").

Overall, participants predominantly mentioned patient information (Case History activity) and test results (Diagnostic Testing activity) in their prioritized lists. Only one or two items from the Patient Chart or Library were mentioned per group. Table 19 presents a summary of the number of items found on participants' prioritized lists. The data show that the residents specified more items overall, and both students and residents included more Case History items than Testing items. A potentially interesting difference between the groups is the fact that students listed roughly half as many Testing items as Case History items, while residents made more use of Testing items and less of Case History items. Experts also consistently used more Testing items than Case History items. Expert 4 did not complete a list for Case 3.

Table 19.

Total Number of List Items (Case History, Testing) by Case and Group

	Case #	Case History Items	Testing Items	Total
Students (n = 10)	1	42	23	
	2	35	19	
	3	46	20	
Total		123	62	185
Residents (n = 10)	1	32	30	
	2	39	17	
	3	39	41	
Total		110	88	198
Experts (n=2)	1	3	12	
	2	3	11	
	3	3*	4*	
Total		10	27	36

* Expert 4 did not fill out a Prioritized List for Case 3.

Table 20 provides a breakdown of the number of different items included on the Prioritized Lists of students, residents, and experts. It shows that students also listed a wider range of Case History items than residents did for two out of three cases and overall. At the same time, residents listed a wider range of testing items than students did in two out of three cases and overall. Experts listed a wider range of Testing items than Case History items. Expert 4 did not complete a list for Case 3.

Table 20.

Number of Different List Items (Case History, Testing) by Case & Group

	Case #	Case History Items	Testing Items	Total
Students (n = 10)	1	18	13	
	2	10	11	
	3	15	9	
Total		43	33	76
Residents (n = 10)	1	17	21	
	2	8	8	
	3	12	18	
Total		37	47	84
Experts (n=2)	1	5	11	
	2	2	11	
	3	3*	4*	
Total		10	26	36

* Expert 4 did not fill out a Prioritized List for Case 3.

Ranking of Prioritized List Items

The ranked lists vary in the number of items they contain, since the participants were able to include any number of items in their lists. The analysis will focus on the highest ranked items. Table 21 presents a breakdown of the items ranked highest (1 and 2) on the lists of students and residents. A number of different pieces of information were highly ranked on the lists. For most cases, residents listed more Testing items than the students did.

Case 1: Celiac (difficult). Students' highest rankings were predominantly applied to Case History information (18), and a few high ranks were assigned to Testing information (2). The residents' highest ranked list items included both Case History (10) and Testing items (10). Several students and residents ranked the first data item, diarrhea, as most important. The experts also ranked this item as important, as shown in Table 22.

The results also show that the residents more frequently ranked test results as first or second in importance.

Case 2: Shigellosis (medium). For Case 2, students' and residents' highest ranking, "1", was assigned almost entirely to Case History information. Further, residents grouped the Case History information ranked "1" more frequently. Experts included more Testing information (6 items) than Case History items (3), although both types were ranked "1" and "2" in their lists.

Case 3: Hepatitis A (easy). For Case 3, a number of students (8) and residents (7) assigned the same Testing item (the positive result for the Hepatitis A titer) the highest rank on their prioritized lists. Unlike the other cases, this case included a test result that was definitive and most participants ranked it highly. Expert 4 did not complete the prioritized list for this case. Expert 3 also assigned the Hepatitis A titer (Testing) the highest ranking, along with both another Testing item and a Case History item. Additional items of both types were also listed.

Table 21.

Prioritized List Items Ranked 1 and 2 by Group, Case, and Data Type

	Number of List Items Ranked "1" or "2"					
	Case 1: Celiac (difficult)		Case 2: Shigellosis (medium)		Case 3: Hepatitis A (easy)	
Group	Case History	Testing	Case History	Testing	Case History	Testing
Students	18	2	19	1	12	8
Residents	10	10	15	5	13	7
Total	28	12	34	6	25	15

Experts' Prioritized Lists

The experts' prioritized lists are presented in Table 22. Expert 4 did not complete the list for Case 3. Except for the main symptom or complaint (diarrhea for Case 1, and bloody diarrhea for Case 2), the experts' lists do not contain the same information. Finally, the experts and other participants frequently grouped several items together, suggesting that they were thinking of a more complex clinical picture rather than a list of discrete items.

It should also be noted that Expert 3 expressed concern about how this activity was conducted. To summarize, she indicated that different pieces of information were important for different reasons. She questioned the idea of specifying which information was important for making a diagnosis.

Table 22.

Experts' Prioritized Lists

Expert	Case 1: Celiac (difficult)	Case 2: Shigellosis (medium)	Case 3: Hepatitis A (easy)
E3	1) chronic diarrhea 2) young male 3) no blood 4) anemic, elevated MCV 5) decreased iron, folate & carotene 6) normal GI workup	1) signs of dehydration (increased Hg & Ng, BUN/Creatinine) 2) bloody diarrhea 3) acidotic, decreased $HC0_3$	1) jaundice & elevated bilrubin 2) anemia, fatigue, & weakness 3) liver function tests
E4	1) 20 lb weight loss 2) intermittent/ nocturnal diarrhea 3) HIV negative 4) laxatives negative 5) endoscopies negative 6) stool parasites negative	1) 4 days bloody diarrhea 2) absence of travel or contact 3) Clostridium Difficile negative 4) stool negative	(didn't complete)

Expert Trace

The participants responded to the Expert Trace information in a variety of ways. For this activity, there was no written task, so the verbal protocols are the sole source of data for this analysis. Participants' responses are summarized, and a few examples are provided.

Table 23 presents a breakdown of the types of responses that participants made to the Expert Trace for each case. Briefly, the responses are divided into "No Response"(participant did not make any verbalization concerning the Expert Trace information), "Read Only" (participant read the Expert Trace information aloud but did not make any further comments), and "Comment" (participant made some comment on the information in the Expert Trace). In many instances the participants either made no response or read the Expert Trace without comment.

No response and read only. The responses in these categories cannot be examined further. Most student responses (23/30) fell into these two categories. The resident responses falling into these categories were also fairly high (18/30). For the experts, only one protocol had no response.

Comments. Some examples of the participant comments are listed in Table 24.

Table 23.

Breakdown of Responses to the Expert Trace by Case and Group

	Case 1: Celiac (difficult)		Case 2: Shigellosis (medium)		Case 3: Hepatitis A (easy)	
Response Type	Students (n = 10)	Residents (n = 10)	Students (n = 10)	Residents (n = 10)	Students (n = 10)	Residents (n = 10)
No Response	1	-	3	3	4	1
Read Only	5	4	6	5	4	5
Comments	4	6	1	2	2	4

Table 24.

Expert List: Comment Types and Corresponding Examples

Type	Example
1) Query Expert	1. S6 (Case 1): "*Maybe (the expert) thought of hemochromatosis.*" 2. R8 (Case 1): "*(The expert is more focused on) lifestyle, stress, and anxiety...He jumped probably more into organic disease...*" 3. R1 (Case 3): "*Maybe (the expert) thought Hep A titer is the first one, and the history of travel is important.*" 4. S9 (Case 2): "*Ok so (the expert is) leaning towards fever, which is important...So again these are important but I don't see where this is leading.*"
2) Evaluate List Item	1. S8 (Case 1): "*Uh just weight loss but not a change of appetite um which is definitely very, very important.*" 2. R10 (Case 1): "*Good idea to do an HIV on anyone (with) weight loss.*" 3. R5 (Case 3): "*(The Hep A titer) is most important for diagnosis, but I'm not sure if you want (it only for diagnosis).*"
3) Evaluate Self	1. S2 (Case 1): "*Uh chronic history...I'm getting too confident and missing everything.*" 2. R1 (Case 1): "*I said thoroughly investigated but I missed to talk about HIV.*" 3. R5 (Case 1): "*That's true, HIV would have been (useful).*" 4. R6 (Case 1): "*I didn't ask about his appetite with the weight loss.*" 5. S10 (Case 3): "*(The expert is) obviously smarter than I was.*"

As shown in the table, the comments from all participants roughly fall into three categories:

1. *Query Expert*: the participant refers to what the expert was or might have been thinking.
2. *Evaluate List Item*: where the participant evaluates a particular list item.
3. *Evaluate Self*: where the participant evaluates their own list items, usually with reference to the Expert List.

Summary

Generally, the protocols were sparse for the post-problem reflection activities. The Categorizing activity was not completed by all of the participants, and seems problematic with medical personnel. The Prioritizing activity data showed that participants identified primarily case history data and test results as important information. Students included Case History information in their lists more frequently than residents did. For this activity, experts consistently used more Testing information in their prioritized lists. The Expert Trace activity also elicited various comments. These findings do not relate directly to other research, since they are tasks specific to the BioWorld system.

Additional Findings: Analysis of Post-Questionnaire Data

Finally, the results of the Post-Questionnaire are summarized in this section. These analyses do not address a specific research question. Rather, they are reported here in order to characterize the opinions of the participants concerning their experience with BioWorld. They also provide some support for the authenticity of the experimental task. The questionnaire assessed their general opinions concerning the BioWorld activities. The researcher was interested in determining, broadly, if the participants found that working with BioWorld cases was seen as a potentially valuable learning experience by these populations. The experts were not asked to complete this questionnaire.

Ratings for BioWorld Cases & Diagnostic Accuracy

The first set of questions consisted of three items of the following type:

"In your opinion, the BioWorld cases were ________________________ ".
realistic / interesting / difficult

A Likert scale was used by participants to rate the statements (1 = strongly agree, 2 = agree, 3 = neutral, 4 = disagree, and 5 = strongly disagree). The results are reported in Table 25. As the findings indicate, both groups were generally in agreement with statements indicating that BioWorld cases were realistic and interesting (mean ratings between 1.3 and 1.8). The students indicated that they found the cases to be slightly more difficult, with mean ratings of 2.5 as compared to residents at 3.2.

The ratings of individual participants on these items were compared to the diagnostic accuracy ratings given by the expert (see "General Results" at the beginning of this chapter). Student S8, who received the highest mean accuracy rating (3.33) across the three cases, also rated the cases as less difficult ("4") than the average for students. Similarly, the two residents who achieved the highest mean accuracy ratings (R5 at 3.67, and R7 at 4.0) also rated the cases as less difficult than the average for residents (3.2). Both residents also gave ratings of "4" for the level of difficulty.

Table 25.

Post-Questionnaire Results: General Ratings

"BioWorld Cases are…"	"Realistic"	"Interesting"	"Difficult"
Students (n=10)	M = 1.8 SD = 1.34	M = 1.3 SD = 0.45	M = 2.5 SD = 0.84
Residents (n=10)	M = 1.4 SD = 1.01	M = 1.8 SD = 0.42	M = 3.2 SD =0.89

Ratings for BioWorld Activities

Participants were also asked to rate each BioWorld activity in terms of how useful it was for helping them learn. The results are presented in Table 26. The item was worded as follows: "Please rate the major BioWorld Activities in terms of utility for helping you learn." (1 = very useful, 2 = useful, 3 = neutral, 4 = not very useful, 5 = waste of time).

Table 26.

Post-Questionnaire Results: Utility

BioWorld Activity	Students (n=10)	Residents (n=10)
1) Reviewing case information		
Mean	1.78	1.6
SD	0.97	0.52
2) Ordering diagnostic tests		
Mean	2.1	2.1
SD	0.74	0.74
3) Specifying a diagnosis		
Mean	1.6	2.22
SD	0.52	0.97
4) Categorizing evidence		
Mean	2.4	2.5
SD	1.07	0.58
5) Prioritizing evidence		
Mean	1.8	2
SD	1.0	0.82
6) Reviewing expert trace		
Mean	1.7	1.6
SD	1.06	0.7

Other Participant Comments

The Post-Questionnaire also provided the participants with the opportunity to write any additional comments concerning their experiences with the cases. These additional comments are recorded in Table 27. Briefly, both the written comments and some verbal discussions between participants and the researcher (after experimental sessions) identified several suggestions for improving BioWorld activities.

Participants' Recommendations

Participants' recommendations, made both verbally and in writing on the Post-Questionnaire, are relevant for the ongoing development of BioWorld. They are summarized in this section.

1. *Make the Case History more realistic.* Several of the participants and the two experts indicated that more realism was needed in terms of the case history. Briefly, they suggested that the real experiences of doing a case history involve a dynamic search and data collection relevant to a specific patient. BioWorld should also provide practice in active data collection.
2. *Include more detailed patient information.* Several participants also indicated that more information is often required and should be accessible on-demand.
3. *Develop an interactive physical exam.* Similarly, participants noted the lack of an interactive physical exam. As for the case history, the BioWorld user should practice active data collection in the context of a physical exam component.
4. *Refine BioWorld's Belief Meter.* This tool, which is used for specifying one's level of certainty in a posted diagnosis, needs to be refined. Early in the data collection process, several participants reported having difficulty specifying a percentage level of certainty as required by this tool. For this reason, subsequent participants were not required to specify a level of certainty, although some did.
5. *Improve or replace the library.* BioWorld's library was not well-received. It was described as "not very extensive", "not helpful", "limiting", and "potentially... helpful with more diseases and characteristics". One suggestion was made to replace it with a link to an on-line medical library.
6. *Define the diagnostic tests.* The Diagnostic Tests available in BioWorld are not defined in the system. One participant suggested that information about the tests themselves should be available. This may include information such as how reliable they are in detecting particular conditions or diseases.

Table 27.

Additional Participant Comments Regarding BioWorld Activities

Activity	Participant	Comment
Case History	S1	"putting in negatives" (the history should contain pertinent negative information e.g., symptoms not present, as well as pertinent positive information)
	S2	"more information sometimes needed"
	S3	"limited information on patient history"
	S9	"good but might need more relevant information, on demand for example"
Physical Exam	R2	"physical exam should be an important part of this exercise"
Diagnostic Testing	S2	"(numerical) values would be appreciated" and " defining the tests would also help"
	S7	"should be less limiting, more realistic"
	S9	"ok"
	R3	"using ultrasound/CT would be useful"
Categorizing		
Prioritizing	S9	"good"
Expert Trace	S2	"very good"
	S4	"important"
	S4	"crucial"
	S9	"good"
Library	S4	"not helpful"
	R4	"limiting"
	S7	"not very extensive"
	S9	"not very helpful but potentially with more diseases and characteristics"
General	S6	"very good initiative"
	S2	"more information in general would be good"
	R4	"very useful experience"

Summary

As the post-questionnaire results demonstrate, the students and residents provided generally positive or supportive responses to most items. They found the activities generally interesting and realistic. Based on the ratings, neither group found the cases difficult, but students rated them as more difficult than residents did. The participants' suggestions are incorporated into the recommendations presented in the following chapter.

Chapter Summary

This chapter has presented and discussed the results of the various analyses conducted in the attempt to answer the research questions for this study. Each research question was addressed individually, and the findings were linked to the relevant literature where appropriate. The analyses identified several characteristics of medical students and residents, and compared their cognitive processing to that of experts. The introductory results (time on task, etc.) have provided a general description of the data. The expert models, the coding scheme and its application, provided a characterization of the cognitive factors involved in problem solving in BioWorld, as well as some indication of how the groups differ in this area. It was also important for examining the use of strategic thinking or planning. These results were also linked to relevant studies of medical expertise. The Post-Problem Reflection activities of the groups were also examined. Finally, the Post-Questionnaire data and other participant comments yielded some general opinions about BioWorld cases and activities, and some further suggestions for the future improvement of the system. In the following chapter, the results are used to draw implications for the ongoing development of the BioWorld system for undergraduate medical education.

CHAPTER 5: EDUCATIONAL IMPLICATIONS: THE DESIGN OF BIOWORLD

Introduction

As indicated in the introductory chapter, the second general goal of this study was to apply the results to the ongoing development of the BioWorld environment. This chapter therefore provides a set of empirically-based suggestions for modifying the current prototype specifically for medical education. Several of the recommendations actually deal with ways to extend BioWorld's design for undergraduate medical education. Ideas for some new system features, based on the results of this study and other relevant research, are also provided.

BioWorld already is a promising learning environment for undergraduate medicine, as evidenced by the participants' general engagement in the tasks, their comments on the Post-Questionnaire, and the experts' enthusiasm in contributing to this study. The contents of this chapter should be viewed as potential improvements, rather than as a coherent design plan.

Potential Future Modifications for BioWorld

Supporting Complex Cognition as Problem Solving Takes Place

Active Learning

BioWorld is designed to support active problem solving; learners engage in both diagnostic problem solving and post-problem reflection. However, the new populations have different characteristics as compared to high school Biology students (the original BioWorld population). The design could be extended so that learners focus on important aspects of their processing during problem solving in accordance with the current findings. For example, the *Select Hypothesis* menu could be modified to allow for listing multiple hypotheses and their associated probabilities, and the ongoing modification of hypotheses as they proceed with a case.

Developments in this area should also incorporate other research on instructional strategies and models, including PBL and tutoring in medicine (Custers et al., 2000; du Boulay & Luckin, 2001; Evenson & Hmelo, 2000; Hmelo, Gotterer, & Bransford, 1997; Jonassen, 1996; Lajoie, et al., 2001; Norman, Brooks, Colle, & Hatala, 1999; Yip, 2002).

Adaptivity and Learner Modeling

The current BioWorld version provides basic feedback messages (e.g., whether a diagnostic hypothesis is correct or not). There are many options for developing more adaptivity in the system, e.g., if some of the modeling and assessment components are developed, they will provide the basis for making instructional adaptations based on learner performance. The BioWorld system would ideally be usable by medical students at different points in their training. In this case, the kinds of feedback and help should then change with the learner's experience. The Case Builder authoring tool currently being developed might provide options for adapting cases to individual users or particular levels of expertise. The case author could design feedback or case information to be presented to learners as they solve a case (see *Authoring*).

To be able to interpret an individual's problem solving actions, BioWorld must have some means for accessing the reasoning behind them. For example, if a learner is collecting additional case history information, BioWorld can only "know" why s/he is doing this if it somehow asks the learner to specify the reason for doing so. How the system should respond also has to be considered. The findings from this study suggest some interesting learner modeling options. For example, the system may be designed to track the types of plans a learner generates, how efficiently s/he progresses through a problem, the kinds of diagnostic hypotheses entertained at different points during problem solving, how new data is interpreted, and even the learner's self-evaluative and self-regulating actions. This information could be very useful for both assessment and feedback for the learner.

Assessment

Problem-based assessment. Both educational and medical researchers have examined issues concerning the measurement of problem solving. The need for assessing learning in problem-based medical settings has been repeatedly expressed (Charlin, Brailovsky et al., 2000). The types of data that BioWorld collects and uses to support learning will probably have to be revised to incorporate the results of the study. Some general guidelines may be derived from the medical education literature. For example, Hannan, Volkan, Fishman, Silvestri, Simon, and Fletcher (2002) present the *Objective*

Structured Clinical Examination (OSCE), which has been used to assess the ability of medical students to perform various types of diagnostic tasks.

Researchers have long noted the need for assessments of real patients as a measure of diagnostic skill for medical students, as opposed to the use of standardized multiple-choice tests (Barrows, Williams, & Moy, 1987). Schuwirth, Blackmore, Mom, van den Wildenberg, Stoffers, and van der Vleuten (1999) address case writing in medicine for the purposes of assessment. Other researchers are working on extending the illness script theory in terms of medical education. Charlin, Tardif, and Boshuizen, (2000) discuss the role of illness scripts as providing a theoretical framework for the structure of medical diagnostic knowledge and problem solving. They also consider the implications of script theory for instruction and assessment. Finally, the Script Concordance Test described by Charlin, Brailovsky, et al. (2000) may be useful in designing assessment for BioWorld; it allows for the assessment of a medical student's illness script in relation to the scripts of experts. It also includes a scoring system that assesses knowledge organization and use.

Focusing on computer-based environments, Baker and O'Neil (2002) examined the distinction between domain-dependent (e.g., subject matter) and domain-independent knowledge (general strategies used across domains), and authoring issues for assessment. Decisions about the assessment of learning through problem solving should strike a balance between these two types of knowledge depending on the context. Finally, it would be valuable to explore dynamic assessment possibilities (Lajoie, 2003) for BioWorld. Briefly, if a learner model is collecting data as a learner progresses through a problem, this gives the potential for creating feedback for the learner right in the problem solving context.

Making Cognitive Processes Explicit

Various researchers have already expressed the need for medical training to help students to make certain "generic" components of their thinking explicit to help them focus on their own thinking and learn how to conduct diagnostic tasks (e.g., Custers et al., 2000; Lajoie, et al., 1998). From AI in medicine perspective, Magnani (1992) suggests that computer-based systems should be used that allow students to explicitly distinguish between medical knowledge and reasoning processes used during problem solving.

At a general level, the learner can be supported while developing an explicit, generic model of the main steps or processes to be carried out during diagnosis. At a lower or more local level, support for specific situations encountered during problem solving can also be provided. Following are some possible ways that thinking can be made explicit to aid learning in BioWorld.

Planning - Collecting Data - Hypothesizing

The expert models revealed an apparently systematic cycle of planning, collecting subsequent data, and hypothesizing by experts. The comparative lack of planning by the other groups suggested that these groups did not proceed in this same manner. Students engaged in comparatively little planning, while residents more closely resembled experts, both in frequency and content. This suggests that students may benefit from assistance to develop their planning skills and practice using them in the context of realistic cases. Also, the amount of planning was apparently influenced by the difficulty or features of cases.

There are a number of possibilities for modifying BioWorld based on these findings. A planning tool could be developed that can elicit plans from students, and support them as they modify plans (e.g., in the face of new data), evaluate plans, and subsequently assess whether plans were carried out. Support for developing various types of plans, including general plans for proceeding with a case, and more specific plans for testing to rule in or rule out specific diagnostic hypotheses, would also be helpful. Planning in response to particular features of a case may perhaps be supported by allowing for flexibility in the depth-of-planning as well as emphasis on the appropriate amount of type of planning. For example, BioWorld could help students develop and complete plan(s) in order to choose between competing hypotheses. By working on explicitly represented plans and associated hypotheses, some carry over to actual diagnostic situations should occur.

Apparently, little research in medical settings deals with tools to support planning in the context of diagnostic problem solving. The SICUN (see Lillihaug & Lajoie, 1998) is one example where, with the help of a Reflection Graph, users can reflect on their plans and actions in the context of diagnoses. CPA or Clinical Problem Analysis (Custers et al.,

2000) is a comprehensive approach to diagnosis and a teaching method, consisting of several generic steps for medical diagnosis that include developing an action plan.

Some work outside of the medical domain is also relevant. For example, Miller (2001) describes a cognitive approach for the development of tools that support planning in the context of military air campaigns. Ge and Land (2003) investigated two scaffolding strategies: question prompts and peer interaction, in an ill-structured task. The prompts encourage students to "attend to important aspects of a problem at different phases and assist them to plan, monitor, and evaluate their solution process" (p.24). Students who received the prompts apparently performed significantly better than those who did not.

Linking and Evaluating

Experts also made more use of linking and evaluating operators than did the other participants. This supports the general view of expert knowledge as richly integrated. The experts in this study apparently saw more links between the discrete items of patient information. An activity or tool could be developed to encourage learners to develop links among available data. They could practice reviewing and examining discrete data items in conjunction with others, perhaps with the help of feedback and/or hints authored by a medical expert or teacher. For example, if an important link between two pieces of data exists, the author might set up a hint to be activated once one of the pieces of information is accessed. The hint might take the form of a question (e.g., "What other information is important to consider in conjunction with this?"). In terms of evaluating, a case author may also decide in advance which items warrant careful examination, and design hints and feedback to encourage it.

Initial Review of Case Information (and Physical Examination Results)

Another finding of interest was the fact that experts spent much effort reviewing case history information, apparently exploring and developing their understanding of the problem. They identified the correct diagnostic hypotheses early in the diagnostic process, and developed plans in response to data. BioWorld should therefore provide support for the careful review of the case history and physical examination, and the use of this information for generating hypotheses and plans. Learners would benefit from support for data collection and evaluation, developing a set of hypotheses, and proceeding with diagnostic testing. Clearly, this could also link in with the recommended planning tool, so that students learn to systematically plan, collect data, evaluate their results, and plan their next action(s).

Self-Knowledge, Performance, and Regulation

The use of knowledge states relating to *Self* (Knowledge, Performance, and Regulation) was comparatively low for all groups, but they demonstrate that participants did change their focus from the problem to their own performance of knowledge at multiple points. Two basic kinds of modifications could be considered for BioWorld. First, an evaluative tool that a learner uses to examine their own knowledge or performance might fit very well with the Post-Problem Reflection activities. It may also be useful to consider the possibilities for authoring in conjunction with self-evaluation during the *Problem Solving* phase. For example, a case author might "program" self-evaluation activities into the case (e.g., prompts to evaluate problem solving success).

Second, the use of self-regulation knowledge states also suggests that a self-regulation tool might be useful. It could explicitly present a "generic" model for diagnosis, and prompt learners to consider their progress in relation to this model. In this way, they might develop more systematic self-regulation skills. Zimmerman & Campillo (2003) present their cyclical three-phase model of self-regulatory activities that could fit well with the BioWorld phases: 1) fore-thought, 2) performance, and 3) self-reflection. A lot of other potentially valuable self-regulation research exists (Ertmer & Newby, 1996; Ley & Young, 2001; Pintrich, 2000; Winne, 1997; Winne & Jamieson-Noel, 2003). Other recent research has examined self-regulation in medical education. For example, Evensen, and colleagues (Evensen & Hmelo, 2000; Evensen, Salisbury-Glennon, & Glenn, 2001) have researched a situated model of self-regulation in a problem-based learning (PBL) curriculum. Henderson and Johnson (2002) developed an email exercise to develop reflection skills in undergraduate medical school. Other related work in PBL should also be consulted (e.g., Abate, Meyer-Stout, Stamatakis, Gannett, Dunsworth, & Nardi, 2000; Friedman & Deek, 2002; Yip, 2002).

Infusing Expert Thinking into the Learning Activities

Access to expertise is an important factor in the development of medical students' expertise as diagnosticians. Medical training has a long apprenticeship tradition and many learning experiences in medical school are organized to provide novices with exposure to the thinking of more expert personnel including residents and physicians (e.g., medical rounds, PBL groups).

Differences between the cognitive processing of experts and the other groups appear early in the problem solving process (e.g., the experts placed more emphasis on working through the case history). For example, the experts put a lot of effort into working through the case histories, always generated the correct diagnosis early in the problem solving process, and repeatedly developed plans for how to proceed through cases. BioWorld could engage students in similar types of processing by providing them with appropriate tools and/or activities. Access to expert thinking could be provided at different points during problem solving, and may include exposing them to examples from experts doing the same tasks. In BioWorld's current prototype, learners only have limited access to expert reasoning during the last post-problem reflection activity, *Expert Trace*. They could benefit from different types of exposure to expert thinking at various points during problem solving. For example, if a learner has worked through a case history and is ready to proceed with a physical examination, it may be effective for a simulated expert to question his or her understanding of the case and perhaps provide feedback or an expert summary at that point.

Another obvious possibility for infusing expert thinking may be to present expert plans at particular points during problem solving. If at all possible, how expert plans and hypotheses change as a function of new information would also be very valuable for learning. Other options include a panic button to access expert feedback, recaps and reviews of planning steps and the actions carried out to achieve them, and constrained hints based on expert performance.

Authoring

Since the time that this study was conducted, work has begun on developing BioWorld's Case Builder (Lajoie, 2003). It provides basic authoring capabilities, but new capabilities may be added based on these results. Clearly, if medical teachers are going to adopt BioWorld for training their students, they will want the system to have some basic capabilities. Specifically, they will want to be able to create cases that have the features they want. For example, they may also want to be able to create feedback, hints, etc. to help students with links, plans, evaluation of evidence, etc.

Promoting Authenticity, Realism, or "Situatedness"

The general principle of authenticity in learning activities has much support in the Educational Psychology literature (see Adelson, 2003). This principle underlies BioWorld's original design; it is designed to engage learners in "realistic" problem solving. For the undergraduate medical population, realism is clearly important for all aspects of the cases and activities. Future development should therefore include modifications to both 1) the content and structure of cases, as well as 2) the types of support the system provides as a learner is attempting to solve a case. The results of this study point to several ways to improve the level of "realism" of BioWorld activities.

This realism should involve incorporating realistic tools and procedures from the appropriate medical sub-domains. Currently, the BioWorld system can present cases from virtually any non-visual medical domain. The features of the particular sub-domains should be examined and incorporated into BioWorld.

Problem Solving Activities

Case history. A general recommendation for BioWorld's case history activity originates from the participants' opinions as indicated in their responses to the Post-Questionnaire items and in conversation after the data collection sessions. The Patient Chart (Case History) would be more "realistic" or authentic if it allowed learners to collect additional data about a patient. In fact, an important component skill of medical diagnosis is the ability to search for data efficiently and effectively. BioWorld's diagnostic testing component is already interactive in this way.

As mentioned previously, Patel, et al. (1997) highlighted the importance of the history taking and the fact that experts use this information to develop rich, integrated knowledge structures that they use throughout the problem solving process. In contrast, less-experienced individuals do not create the same quality of knowledge structures. It seems very important then to model the "realistic" learning situation by providing this sort of interactivity for history taking activities. This would entail restructuring the activities such that the learner would be able to actually "take" a case history, and then "conduct" a (focused) physical examination. Iit will be important to draw on the established, hierarchically-organized procedures for taking a case history that are taught in medical schools (see Billings & Stoeckle, 1999) and routinely practiced by physicians.

These procedures should serve as the structured knowledge base for the BioWorld activities, such that taking a case history or performing a physical exam would follow the same general procedures. A pull-down menu system would be well-suited as these procedures are already hierarchically-organized (in a checklist format). The learner could then proceed with a case history by looking at the higher level categories or major topics such as past medical history. The learner could then select the specific items they want more information about from the menu. The learner would then control how, when, and what information was collected. This interactive model would more closely resemble how diagnostic activities proceed in real-life medical settings.

Physical exam. The Patient Chart currently in BioWorld could be maintained, as physicians often receive this sort of information prior to questioning or examining a patient. However, as for the case history, participants also identified the need for interactivity for the physical examination. The crucial modification to this activity would be a simulated physical examination. Similar to taking a case history, established procedures exist for the physical exam, and they are also organized in the same kind of hierarchical format (for an example, see Fraser, 1987). BioWorld's physical examination could also be arranged in this way.

Diagnostic testing. As mentioned, BioWorld's diagnostic testing activity already is interactive. This is an important strength of the system. Learners make testing decisions in a manner similar to how it is done in medical practice. That is, they decide what tests to order, what sequence to order them in, how to interpret any test results, etc. However, there are certain modifications that would be helpful in making the testing component more realistic. First, BioWorld currently presents all available test names in a list and returns the test results immediately once they are requested. Based on protocol excerpts and specific comments from participants, there are actually a number of factors that influence testing decisions that are not addressed in BioWorld. For example, different tests take different amounts of time to conduct (some take weeks to receive results). There are also differences in the "costs" associated with various tests, including the level of invasiveness or discomfort for the patient, and the actual monetary expense to produce a test result. In addition to their diagnostic utility, these are apparently important influences on how testing is actually conducted.

Values could be assigned to each test in BioWorld to reflect some of these considerations or "costs". For instance, each test might have a time delay for results to be received, to help the problem solving process unfold in a more realistic sequence. If in actual practice a particular test result is typically received one month after it is ordered, an alternative testing plan may be adopted. This should also happen in BioWorld. Not that learners in this context should literally wait, but they should not simply access a test result that is associated with a long delay before one that is not. The level of invasiveness and expense should also be considered; physicians will try to not use test that is highly invasive or expensive if other means can be used instead. The same considerations should be part of the testing process in BioWorld.

New Learning Activities

Case presentation. If the case history and physical examination activities are modified so that learners collect data themselves, it will be desirable for BioWorld to help them focus and organize their data. Again, learning to work systematically with case history and physical examination data is apparently important based on the results of the study. One possible addition to this part of the system would be a *case presentation* option, which could be presented after the case history, after the physical examination, or after diagnostic testing. That is, once a learner has collected all the data that s/he thinks is relevant for a particular case, the information should be organized and used to focus further investigations. The ability to succinctly present the relevant information about a particular patient is an important medical skill (Lajoie, et al., 2001) that is practiced repeatedly during medical training.

Treatment. Solving cases in BioWorld currently does not include working out treatment options. This is obviously a critical part of medical practice, and it could be a very important improvement for BioWorld. During the course of this study, some participants spontaneously mentioned treatment options for the cases. With a simulated treatment activity, learners could perhaps administer treatments to a patient and see the results. They could also be exposed to the treatment plans of experienced physicians.

Different learning roles and collaboration. Another possibility for BioWorld would be the design of different learning roles. A number of possibilities exist, including group work in medical classrooms, and at a distance via the Internet. For example PBL groups are organized around different roles performed by different students. BioWorld could also provide such roles, perhaps having other roles performed by simulated on-line students and experts. For reviews of PBL learning effects see Dochy, Segers, Van den Bossche, and Gijbels (2003), and Friedman, & Deek, 2002). For research on tutoring strategies in medicine and PBL, see Lajoie et al. (2001). In addition, work on the use of technology in PBL groups should also be consulted (Arts, Gijselaers, & Segers, 2002; Orrill, 2002). Some tools have already been developed that incorporate PBL type activities (Abate, et al., 2000; Shortliffe, Barnett, Cimino, Greenes, Huff, & Patel, 1996; Yip, 2002). In any event, the collaborative possibilities would be worth considering, if only to aid in the ongoing design of tools and any simulated personnel that may be built into the system (e.g., experts).

Complex, Realistic Patients

The fact that real patients often have multiple medical problems that require attention is also important from a design perspective. Researchers in medical problem solving have typically used cases that are not realistic in this respect (Custers et al., 2000). Complex patients would have implications for learning activities and authoring. In terms of learning activities, learners need to have support as they diagnose a specific problem and sort out a patient's additional problems. Incorporating the Problem List activity, which is an integral part of the diagnostic process (see Wiseman, 2004), would be a major improvement. Briefly, the problem list is a set of medically relevant patient problems derived from examination. They consist both the active problem that requires diagnosis and any other significant conditions or previous medical experiences. The *Problem List Generator (PLG)* is an example of a computer-based tool designed to elicit a problem list from learners (Danielson et al., 2003). Finally, implications for the authoring are to 1) allow for a patient to have multiple, sometimes conflicting conditions, 2) support the learner in distinguishing between more and less relevant problems and which problems require medical attention.

Settings for BioWorld's Use

There are several possibilities for the BioWorld system to be used in medical training contexts, each with accompanying requirements. For example, it may be used by students working alone at home, or by medical teachers who create cases for their students to solve collaboratively in PBL settings. Each potential setting has specific design implications, that require attention (see Kreijns, et al., 2003). If it will be used in medical classrooms, as noted by Anderson, Corbett, Koedinger, and Pelletier (1995), successful integration of a CBLE into a classroom setting depends on addressing certain issues. Some will be less of a concern for introducing BioWorld into medical classrooms than they would be for high school Biology. Medical students are more sophisticated learners, who would see the relevance of the activity for their training. Basics such as skill in computer use on the part of the teachers and learners should also not be a major concern. However, it will be important to "fit" BioWorld into an undergraduate medical curriculum.

Different Modes of Operation

It could be very valuable for BioWorld to have different operational modes, perhaps for learning and testing. In *Learning* mode, all of the tools and supports for learning could be accessible. In *Testing* mode, access to these supports might be removed, and the system may monitor the actions of the learner. Other alternatives include a number of learning modes. For example, Alexe and Gecsei (1996) describe several learning modes in the context of an SICU tutor. In *demonstration mode*, the learner can view a task being conducted in non-interactive or interactive formats. In *Task Resolution* mode, the learner is asked to perform a simulation of the task with coaching.

Promoting Post-Problem Reflection Focused on Complex Cognition

In keeping with the preceding suggestions, post-problem reflection activities should be focused on specific cognitive processes as well as the content of individual cases. The results of the study suggest that the focus of learners' reflection should be expanded from the relative importance of specific data items for making a diagnosis to the components of thinking that are made explicit during problem solving (e.g., planning, linking, evaluating).

Various researchers have investigated the use of reflection in training medical students and their findings may be useful for BioWorld (see Hollen, 2002; Henderson, Berlin, Freeman, & Fuller, 2002). As an example, Henderson and Johnson (2002) present a course design, which involves having medical students reflect on workshop exercises they participated in, and write about the types of impact they felt.

Categorizing

This activity was originally included in the BioWorld version that was developed for high school students. This activity is apparently less appropriate for medical populations. These individuals would clearly be very well aware of the basic distinctions between these data types. The results of the study bear this out, in that participants commented that the activity was not that helpful and often opted not to complete it. So, this activity should be removed or re-worked for this population. It could instead be organized around the various learning activities, and focus on plans, hypotheses, and the use of data to rule in and rule out various diagnostic hypotheses. In this way, the learner could review the steps taken to solve the problem, focusing on more salient information.

It could also be helpful to consider an activity that would help learners to organize their thinking in a way that more closely resembles real medical thinking. The Problem List activity (see Wiseman, 2004) would be one possibility for structuring both the problem solving activities, and part of the post-problem reflection activities. That is, if the problem solving involved building a problem list and using it to diagnose the case, the list could also be used to reflect back on the important and less important aspects of the case. It could also be used to evaluate or critique the problem solving process.

Prioritizing

This activity was designed to focus learner attention specifically on evaluating the data used to make a diagnosis. Again, for high school students this has been an effective learning activity. In this study, individual participants commented that simply rating pieces of information from most to least important is not necessarily that useful. This activity could be re-worked to allow learners to examine their problem solving, perhaps evaluating their hypotheses based on data at different points during problem solving. They may be more useful to reflect on which data raise or lower the likelihood of different hypotheses.

It would be interesting to further investigate this apparent difference between the groups in the type of information they found important to solving the cases. These differences may reflect the residents' greater experience with diagnostic testing. In particular, it would be important to investigate why the participants find particular pieces of information important. In the current form, BioWorld does not elicit the thoughts behind why particular information was selected.

Expert Trace

BioWorld's *Expert Trace* was conceived of as a means for learners to access expertise and use it to reflect on their own performance. It should also be mentioned that the original design of BioWorld included an audio clip of an expert discussing how he solved the case. This component was not functioning at the time of this study and would no doubt have made a difference to the *Expert Trace* activity. However, there are some additional considerations, based on the expert's short lists of items, the generally scant protocols for this learning activity, and some comments from the participants. First, some participants had trouble with the idea of "what information was most important" for solving a case. Some explained that 1) case information can be useful for different things (patient management, diagnosis), 2) they do not typically think in terms of prioritized lists of pertinent information, and 3) more additional information should be available about the expert lists (which would have been partly alleviated by the audio clip).

The *Consult* button currently does allow learners to access basic instructions (e.g., what to do next), but does not make expert thinking explicit for learners. It could be useful to replace this feature with access to an expert's thinking *during problem solving*. This could be accomplished by having audio or videotaped excerpts of an expert working on the same task available to the learners. An end-of-case review of expert problem solving can still be very valuable. This needs to be developed so that more information is available and the experts' solution paths and accompanying reasoning become available.The *Expert Trace* activity could be very useful if the expert's processing is made explicit also. The idea of comparing one's own thinking during problem solving with that of an expert is supported by the data. If a learner has focused on their own planning and other processing, they may benefit greatly from reviewing an expert's summary that also contains these elements. During post-problem reflection, a detailed recounting of the case by an expert may also be very valuable.

Chapter Summary

These general suggestions, based on the results of the study, should be valuable in guiding the ongoing development of the system. Clearly, there are many options for the future development of BioWorld for medical training contexts. This chapter has also highlighted related research that should also be useful for BioWorld's future development. The following chapter presents the conclusions of this study.

CHAPTER 6: CONCLUSIONS

This final chapter presents a summary of what this study has achieved, the potential additions to the literature, and as well as the limitations of the study.

What This Study Has Achieved

This study included a review of the medical expertise literature that highlighted a set of established findings. It also demonstrated that this research has examined a range of cognitive structures and processes, many of which: have ultimately provided similar descriptions of cognitive processing at various levels of expertise, and proved reasonably compatible with the complex cognition involved in medical problem solving. The review also highlighted certain methodological problems that are only sometimes acknowledged in the medical expertise literature, but should be addressed in the interest of extending the existing research on medical problem solving and informing medical education. Briefly, studies of more realistic diagnostic problem solving are required.

The review also identified a general lack of implications for medical training, which points yet again to a need for other types of research that can inform the development of medical training. Some potentially informative research from educational psychology, medical education, and educational computing was identified. Further, the design of this study was presented as an example for merging research on medical problem solving with the development of student-centered learning environments.

This study has provided characterizations of medical students, residents, and experts solving BioWorld cases in internal medicine. Following the BioWorld case format, the experimental task used in the study was more realistic or authentic than many tasks have been in medical expertise studies. Therefore, the problem solving elicited in the study is also somewhat more realistic. In fact, the diagnostic task seems to have elicited some processing that has not been emphasized in the medical expertise literature (e.g., planning). The results of this study included the development and use of a flexible coding scheme that embeds realistic problem solving structures and processes within the BioWorld learning activities. In addition, the cognitive processing of these individuals during diagnostic problem solving was examined in detail, focusing on knowledge structures and processes. As compared to many other studies that were designed to answer more specific research questions (e.g., "Do students and experts use biomedical or clinical knowledge?"), this study provides a detailed examination of cognitive processing. These analyses serve to achieve the first general goal of the study: *to characterize the cognitive processing of medical personnel with varying levels of expertise as they diagnose cases in BioWorld*

It has also demonstrated that CBLEs are a promising direction for medical education, based on the results of this study concerning BioWorld and the varied research studies regarding other CBLEs for medical training. By studying problem solving in a particular learning context, the results of the study can be interpreted and used to draw instructional implications for both the design of this environment and, more generally, for medical education. This study therefore provides a tentative example of empirically-based CBLE design. This information serves to achieve the second general goal of the study: *to use the results of the study to draw instructional implications for the ongoing development of BioWorld as a learning environment.* Many other studies that could potentially inform the future development of BioWorld have been identified.

Potential Additions to the Literature

Chapter 2 provided an overview of the medical expertise literature. The following tentative findings (more research is required) could add to this literature:

- Realistic diagnostic problem solving has often not been what medical expertise research has studied. This has many implications; future work should strive to use more realistic tasks and focus specifically on diagnosis. Further, more realistic tasks will likely result in other factors becoming important for explaining how experts and non-experts actually solve diagnostic problems.
- The coding scheme developed for this study specifies a number of operators and knowledge states that can be flexibly combined. It also allows for identifying and characterizing the use of cognitive structures and processes, as well as examining their use as elements of a learning environment (BioWorld activities).
- The use of N-Vivo as a data management and analysis tool is also potentially of interest to researchers. It allows for the development of coding systems, data coding, and very flexible searching of coded data.
- Planning may play a more important role in diagnostic problem solving than the literature suggests. Based on limited data, there is some evidence that experts consistently make planning statements from early in the problem solving process. Less planning occurred with residents, and even less with medical students. Planning may also be affected by case difficulty, but experts consistently generate multiple plans for cases (easy, medium, and difficult).

- For the first time, medical students and residents have worked on BioWorld cases and given their opinions about the system. According to these groups, the problem solving elicited by the system is already reasonably realistic, interesting, and difficult for these populations. Future research can build on these findings.
- CBLEs for medical training can be developed through a detailed examination of problem solving in particular medical sub-domains, accompanied by a focus on student-centered learning environment design.

Limitations of the Study

This study has several limitations. First, a relatively small number of cases was used (3), all from Internal Medicine. The results therefore are not necessarily generalizable to other sub-domains. Second, the analyses conducted in this study are based on the cognitive approach that focuses on particular aspects of medical diagnosis, but does not address all potentially relevant aspects (e.g., social and environmental factors). In addition, only two experts were available to act as participants. Finally, the BioWorld case format, while more realistic in some aspects than other experimental tasks, has not been used in other studies of medical expertise. As a result of these limitations, the results should be considered as a descriptive, initial account of the cognitive processing by these groups solving BioWorld cases.

Future Research

Based on the results of this study, there are several promising directions for future research. In keeping with the general goals of the study, the results provide basic characterizations of the participant groups and experts. Several directions could be pursued in order to build on these findings. For example, it would be interesting to further examine:

- *different medical populations.* This study resulted in smaller differences between the student and resident groups than expected. In future, students at earlier and later points in their training should be included. This will provide a more representative range of potential BioWorld users.
- *development over time.* Longitudinal studies could be conducted to examine the development of knowledge and skills over the course of medical training. This could involve studying medical students using BioWorld at multiple times in their training.
- *case features* (e.g., difficulty, typicality). The relationship of case features to problem solving (operators and knowledge state use) is important both for

examining problem solving and for designing instruction. How different populations respond to different case features, and what capabilities BioWorld should have as a result are important concerns.

- *linking*. Given that richly elaborated and interconnected knowledge structures are apparently a feature of expertise, a closer examination of the use of linking may be important. Rather than treating data items as discrete, it seems important to both examine how the connections between them emerge as well as how such linking can be supported.
- *the role of planning and goal statements*. Designing more realistic diagnostic tasks, and perhaps including probes designed to elicit plans and goals could provide more detailed information. How planning develops and what kinds of instructional support help the acquisition of planning skills would also be of interest.
- *self-regulation*. Although the data concerning self-regulation were minimal in this study, some differences were observed. This is very interesting from an instructional perspective.
- *post-problem reflection*. Though results were minimal, the groups used different information to solve the cases.
- *new tools and functionalities for BioWorld*. As identified in Chapter 5, many possibilities for new and modified system features exist. Future study could include the design, development, and evaluation of some of these tools by medical personnel.

Summary

This study has highlighted some of the characteristics and weaknesses of medical expertise research. It has also been reasonably effective in identifying some of the characteristics of students', residents', and experts' in diagnostic problem solving in BioWorld. Their problem solving has been contrasted with that of experts. Not surprisingly, the experts differed considerably in several ways from the two less-experienced groups.

This study has also made a case for the development of computer based learning environment based on detailed study of medical problem solving in specific medical sub-domains. The results have also allowed for identifying various potential modifications for the system, which should help to make learning from BioWorld cases more "authentic" and more effective. It has also identified related research that can also potentially inform BioWorld's future development. The contents of this chapter are intended to assist in the future development of the system, so that BioWorld can become a very effective learning and learning environment for medical education.

REFERENCES

Abate, M. A., Meyer-Stout, P. J., Stamatakis, M. K., Gannett, P. M., Dunsworth, T. S., & Nardi, A. H. (2000). Development and evaluation of computerized problem-based learning cases emphasizing basic science concepts. *American Journal of Pharmaceutical Education, 64*(1), 74-82.

Adelson, B. (2003). Bringing considerations of situated action to bear on the paradigm of cognitive modeling: The 2002 Benjamin Franklin Medal in computer and cognitive science presented to Lucy Suchman. *Journal of the Franklin Institute, 340*, 283-292.

Aleven, V., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science, 26*, 147-179.

Alexander, P. A. (2003). The development of expertise: The journey from acclimatization to proficiency. *Educational Researcher, 32*(8), 10-14.

Alexe, C., & Gecsci, J. (1996). A learning environment for the surgical intensive care unit. In C. Frasson, G. Gauthier, & A. Lesgold (Eds.), *Intelligent Tutoring Systems: Third International Conference, ITS '96* (pp. 439-447). Montreal, Canada.

Allen, V. G., Arocha, J. F., & Patel, V. L. (1998). Evaluating evidence against diagnostic hypotheses in clinical decision making by students, residents, and physicians. *International Journal of Medical Informatics, 51*(2-3), 91-105.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4*(2), 167-207.

Anderson, J. R. (1993a). Problem solving and learning. *American Psychologist, 48*(1), 35-44.

Anderson, J. R. (1993b). *Rules of the mind.* Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1991). Is human behavior adaptive? *Behavioral and Brain Sciences, 14*, 471-517.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J. R., Greeno, J. G., Reder, L., M., & Simon, H. A. (2000). Perspectives on learning, thinking, and activity. *Educational Researcher, 29*(4), 11-13.

Anderson, J. R., & Labriere, C. (1998). The *atomic components of thought.* Hillsdale, NJ: Erlbaum.

Anderson, N. (1992). Medical Center: A modular hypermedia approach to program design. In E. Barrett (Ed.), *Sociomedia: Multimedia, hypermedia, and the social construction of knowledge. Technical communication and information systems*, (pp. 369-389). Cambridge, MA: The MIT Press.

Arocha, J. F., & Patel, V. L. (1995a). Construction-integration theory and clinical reasoning. In C. A. Weaver, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch,* (pp. 359-381). Hillsdale, NJ: Erlbaum.

Arocha, J. F., & Patel, V. L. (1995b). Novice diagnostic reasoning in medicine: Accounting for evidence. *The Journal of the Learning Sciences, 4*(4), 355-384.

Arocha, J. F., Patel, V. L., & Patel, Y. (1993). Hypothesis generation and the coordination of theory and evidence in novice diagnostic reasoning. *Medical Decision Making, 13*(3), 198-211.

Arts, J. A. R., Gijselaers, W. H., & Segers, M. S. R. (2002). Cognitive effects of an authentic computer-supported, problem-based learning environment. *Instructional Science, 30*(6), 465-495.

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology, 95*(4), 774-783.

Atkinson, R. M. & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation 2,* (pp. 89-195). NY: Academic Press.

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

Azevedo, R., & Faremo, S. (1999). Expertise differences in radiology: Extending the RadTutor to foster medical students' diagnostic skills. *Paper presented at the 9th International Conference on Artificial Intelligence in Education*, LeMans, France, July.

Baker, E., & O'Neil, H. F. (2002). Measuring problem solving in computer environments: Current and future states. *Computers in Human Behavior, 18,* 609-622.

Barr, A. & Feigenbaum, E. A. (1981). *The handbook of artificial intelligence.* Stanford, CA: HeurisTech Press.

Barrows, H. S. (1996). Problem-based learning in medicine and beyond: A brief overview. *New Directions for Teaching & Learning, 68*, 3-12.

Barrows, H. S., & Feltovich, P. J. (1987). The clinical reasoning process. *Medical Education, 21,* 86-91.

Barrows, H. S., Williams, R. G., & Moy, R. H. (1987). A comprehensive performance-based assessment of fourth-year students' clinical skills. *Journal of Medical Education, 62*, 805-809.

Bartlett, F. C. (1932). *Remembering*. Cambridge, MA: Cambridge University Press.

Bechtel, W., Abrahamsen, A., & Graham, G. (1998). The life of cognitive science. In W. Bechtel, & G. Graham (Eds.). *A companion to cognitive science,* (pp. 3-104). Malden, MA: Blackwell.

Berkow, R. B., & Fletcher, A. J. (1987). *The Merck manual of diagnostic therapy.* Rahway, NJ: Merck & Co.

Best, J. B. (1995). *Cognitive psychology*. Minneapolis, MN: West Publishing.

Billings, A. & Stoeckle, J. (1999). *The clinical encounter: A guide to the medical interview and case presentation* (2nd ed.). Missouri: Mosby Inc.

Bordage, G., & Lemieux, M. (1991). Semantic structures and diagnostic thinking of experts and novices. *Academic Medicine, 66*(9), S70-72.

Bordage, G., & Lemieux, M. (1986). Some cognitive characteristics of medical students with and without diagnostic reasoning difficulties. *Proceedings of the 25th Annual Conference of Research in Medical Education*. New Orleans.

Bordage, G. & Zacks, R. (1984). The structure of medical knowledge in the memories of medical students and general practitioners: Categories and prototypes. *Medical Education, 18,* 406-416.

Boshuizen, H. P. A. (1996). The shock of practice: Effects on clinical reasoning. *Paper presented at the Annual American Educational Research Association.* New York, April 8-14.

Boshuizen, H. P. A. (1994). Cognitive effects of practical experience in high- and low-achieving medical students. *Learning and Instruction, 4*, 313-329.

Boshuizen, H. P. A. & Claessen, H. F. (1982). Problems of research into medical problem solving: Some remarks on theory and method. *Medical Education, 16*(2), 81-7.

Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates, and novices. *Cognitive Science, 16*, 153-184.

Boshuizen, H. P. A., Schmidt, H. G., & Coughlin, L. D. (1987). On-line representation of a clinical case and the development of expertise. *Paper Presented at the Annual Meeting of the American Educational Research Association*, Washington, DC.

Bowdish, B. E., Chauvin, S. W., Kreisman, N., & Britt, M. (2003). Travels towards problem based learning in medical education. *Instructional Science, 31*, 231-253.

Braccio, A. (1988). On-line analysis of novice problem solving in medicine. *Unpublished Master's Thesis,* McGill University, Montreal, Canada.

Bracewell, R. J., & Breuleux, A. (1994). Substance and romance in analyzing think-aloud protocols. In P. Smagorinsky (Ed.), *Speaking about writing: Reflection on research methodology,* (pp. 55-88). Newbury Park, CA: Sage.

Broadbent, D. E. (1958). *Perception and communication.* London: Pergamon Press.

Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). Role of specific similarity in a medical diagnosis task. *Journal of Experimental Psychology: General, 120*(3), 278-287.

Bruer, J. T. (2003). Learning and technology: A view from cognitive science. In H. F. O'Neil & R. S. Perez (Eds.), *Technology applications in education: A learning view,* (pp. 159-172). Mahwah, NJ: Erlbaum.

Bruer, J. T. (1998). Education. In W. Bechtel, & G. Graham (Eds.). *A companion to cognitive science,* (pp. 681-690). Malden, MA: Blackwell.

Buchanan, B. & Shortliffe, E. H. (1984). *Rule-based expert systems: The MYCIN experiments of the Stanford Heuristic Programming Project.* Reading, MA: Addison-Wesley.

Chan, T. W., & Baskin, A. B. (1988). "Studying the prince": The computer as a learning companion, (pp. 194-200). *ITS-88 Proceedings*, Montreal, Canada.

Charlin, B., Brailovsky, C., Roy, L., Goulet, F., & van der Vleuten, C. (2000). The script concordance test: A tool to assess the reflective clinician. *Teaching & Learning in Medicine, 12*(4), 189-195.

Charlin, B., Tardif, J., & Boshuizen, H. P. A. (2000). Scripts and medical diagnostic knowledge: Theory and applications for clinical reasoning instruction and research. *Academic Medicine. 75*(2), 182-190.

Chi, M. T. H. (1997). *Quantifying qualitative analyses of verbal data: A practical guide. The Journal of the Learning Sciences, 6*(3), 271-315.

Chipman, S. F. (2003). Gazing yet again into the silicon chip: The future of computers in education. In H. F. O'Neil & R. S. Perez (Eds.), *Technology applications in education: A learning view,* (pp. 31-54). Mahwah, NJ: Erlbaum.

Cho, K.-L. & Jonassen, D. H. (2002). The effects of argumentation scaffolds on argumentation and problem solving. *Educational Technology Research and Development, 50*(3), 5-23.

Cimino, J. J. (1999). Development of expertise in medical practice. In R. J. Sternberg, & J. A. Horvath (Eds.), *Tacit knowledge in professional practice: Researcher and practitioner perspectives,* (pp. 101-120). Mahwah, NJ: Erlbaum.

Claessen, H. F. & Boshuizen, H. P. A. (1985). Recall of medical information by students and doctors. *Medical Education, 19*, 61-67.

Clancey, W. (1988). Acquiring, representing, and evaluating a competence model of diagnostic strategy. In M. T. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise*, (pp. 343-418). Hillsdale, NJ: Erlbaum.

Clancey, W. J. (1986). Qualitative student models. *Annual Review of Computer Science, 1*, 381-450.

Clancey, W. J. (1984a). Methodology for building an intelligent tutoring system. In W. Kintsch, P. G. Polson, & J. R. Miller (Eds.), *Methods and tactics in cognitive science,* (pp. 51-83). Hillsdale, NJ: Erlbaum.

Clancey, W. J. (1984b). NEOMYCIN: Re-configuring a rule-based expert system for application to teaching. In W. J. Clancey, & E. H. Shortliffe (Eds.), *Readings in medical artificial intelligence: The first decade*, (pp. 261-381). Reading, MA: Addison-Wesley.

Clark, A. (1998). Embodied, situated, and distributed cognition. In In W. Bechtel, & G. Graham (Eds.), *A companion to cognitive science,* (pp. 506-517). Malden, MA: Blackwell.

Cobb, P., & Bowers, J. (1999). Cognitive and situated learning perspectives in theory and practice. *Educational Researcher, 28*(2), 4-15.

Coderre, S., Mandin, H., Harasym, P. H., & Fick, G. H. (2003). Diagnostic reasoning strategies and diagnostic success. *Medical Education, 37*, 695-703.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. In L. B. Resnick (Ed.), *Knowing, learning and instruction: Essays in honor of Robert Glaser,* (pp. 393-452). Hillsdale, NJ: Erlbaum.

Coughlin, L. D. & Patel, V. L. (1986). Text comprehension and expertise in the domain of medicine. *Paper presented at the Annual Meeting of the American Educational Research Association,* San Francisco, April 16-20.

Craik, F. I. M. & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.

Crowley, R., Medvedeva, O., & Jukic, D. (2003). A model-tracing Intelligent Tutoring System for teaching microscopic diagnosis. *Proceedings of the 11th International Conference on Artificial Intelligence in Education*, Sydney, Australia.

C.T.G.V. (2003). Connecting learning theory and instructional practice: Leveraging some powerful affordances of technology. In H. F. O'Neil, & R. S. Perez (Eds.), *Technology applications in education: A learning view,* (pp. 173-209). Mahwah, NJ: Erlbaum.

C.T.G.V. (1996). Looking at technology in context: A framework for understanding technology and education research. In D. C. Berliner, & R. C. Calfee (Eds.), *Handbook of educational psychology,* (pp. 807-840). NY: MacMillan.

Custers, E. J. F. M. (1995). *The development and function of illness scripts: Studies on the structure of medical diagnostic knowledge.* Maastricht, Netherlands: University of Maastricht.

Custers, E. J. F. M., Boshuizen, H. P. A, & Schmidt, H. G. (1998). The role of illness scripts in the development of medical diagnostic expertise: Results from an interview study. *Cognition & Instruction, 16*(4), *367-398.*

Custers, E. J. F. M., Boshuizen, H. P. A, & Schmidt, H. G. (1996). The influence of medical expertise, case typicality, and illness script component on case processing and disease probability estimates. *Memory & Cognition, 24*(3), 384-399.

Custers, E. J. F. M., Boshuizen, H. P. A, & Schmidt, H. G. (1992). The relationship between medical expertise and the development of illness scripts. *Paper presented at the annual meeting of the American Educational Research Association.* San Francisco, CA.

Custers, E. J. F. M., Stuyt, P. M. J., & De Vries Robbé, P. F. (2000). Clinical Problem Analysis (CPA): A systematic approach to teaching complex medical problem solving. *Academic Medicine, 75*(3), 291-297.

Cuthbert, L., du Boulay, B., Teather, D., Teather, B., Sharples, M. & du Boulay, G. (1999). *Expert/novice differences in diagnostic medical cognition – A review of the literature.* Research Paper: University of Sussex, ISSN 1350-3162.

Danielson, J. A., Bender, H., Mills, E. M., Vermeer, P. J., & Lockee, B. B. (2003). A tool for helping veterinary students learn diagnostic problem solving. *Educational Technology Research and Development, 51*(3), 63-75.

Davidson, J.E. & Sternberg, R. (2003). *The psychology of problem solving.* NY: Cambridge University Press.

De Jong, T., & Van Joolingen, W. (1996). *Discovery learning with computer simulations of conceptual domains.* IST-MEMO-96-02. University of Twente: Netherlands.

Del Soldato, T. (1994). *Motivation in tutoring systems.* Cognitive Science Research Report: University of Sussex.

Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003). Effects of problem-based learning: A meta-analysis. *Learning & Instruction. 13*(5), 533-568.

Du Boulay, B., & Luckin, R. (2001). Modelling human teaching tactics and strategies for tutoring systems. *International Journal of Artificial Intelligence in Education, 12*, 235-256.

Dunbar, K. (1998). Problem solving. In W. Bechtel, & G. Graham (Eds.), *A companion to cognitive science,* (pp. 289-298). Malden, MA: Blackwell.

Elliot, C. E., Williams, K. A., & Wolff, B. P. (1996). An intelligent learning environment for advanced cardiac life support. *Proceedings of the American Medical Informatics Association*, Washington, DC, October, pp. 7-11.

Elstein, A. S. (2001). Naturalistic decision making and clinical judgment. *Journal of Behavioral Decision Making, 14*(5), 363-365.

Elstein, A. S. (2000). Clinical problem solving and decision psychology: Comment on "The Epistemology of Clinical Reasoning". *Academic Medicine, 75*(10), S134-S136.

Elstein, A. S. (1994). What goes around comes around: Return of the hypothetico-deductive strategy. *Teaching and Learning in Medicine, 6*(2), 121-123.

Elstein, A. S., & Rabinowitz, M. (1993). Medical cognition: Research and evaluation. In M. Rabinowitz (Ed.), *Cognitive science foundations of instruction*, (pp. 189-201). Hillsdale, NJ: Erlbaum.

Elstein, A. S., & Schwarz, A. (2002). Clinical problem solving and diagnostic decision making: Selective review of the cognitive literature. *British Medical Journal, 324*(7339), 729-732.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1990). Medical problem solving: A ten-year retrospective. *Evaluation and the Health Professions, 13*(1), 5-36.

Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning*. Cambridge, MA: Harvard University Press.

Ericsson, K. A. (2003). The acquisition of expert performance as problem solving: Construction and modification of mediating mechanisms through deliberate practice. In J. E. Davidson, & R. J. Sternberg (Eds.), *The psychology of problem solving*, (pp. 31-83). NY: Cambridge University Press.

Ericsson, K. A. (1998). Protocol analysis. In W. Bechtel, & G. Graham (Eds.), *A companion to cognitive science,* (pp. 425-432). Malden, MA: Blackwell.

Ericsson, K. A. & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49*(8), 725-747.

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

Ertmer, P. A. & Newby, T. J. (1996). The expert learner: Strategic, self-regulated, and reflective. *Instructional Science, 24,* 1-24.

Evans, D. A. (1989). Issues of cognitive science in medicine. In D. A. Evans & V. L Patel (Eds.), *Cognitive science in medicine: Biomedical modeling,* (pp.1-19). Cambridge, MA: MIT Press.

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

Evensen, D. H., & Hmelo, C. (2000). *Problem-based learning: A research perspective on learning interactions.* Mahwah, NJ: Erlbaum.

Evensen, D. H., Salisbury-Glennon, J. D., & Glenn, J. A (2001). A qualitative study of six medical students in a problem-based curriculum: Toward a situated model of self-regulation. *Journal of Educational Psychology, 93*(4), 659-676.

Feltovich, P.J. & Barrows, H. S. (1984). Issues of generality in medical problem solving. In H. G. Schmidt & M. L. De Volder (Eds.), *Tutorials in problem based learning,* (pp. 128-143). Assen/Maastricht, Netherlands: Van Gorcum.

Feltovich, P. J., Johnson, P. E., Moller, J. H. & Swanson, D. B. (1984). LCS: The role and development of medical knowledge in diagnostic expertise. In W. J. Clancey & E. H. Shortliffe (Eds.), *Readings in medical artificial intelligence,* (pp. 275-319). Reading, MA: Addison-Wesley.

Fraser, R. C. (1987). *Clinical method: A general practice approach.* London: Butterworths.

Frederiksen, C. H. (1975). Representing logical and semantic structure of knowledge acquired from discourse. *Cognitive Psychology, 7,* 371-458.

Friedman, R. S., & Deek, F. P. (2002). Problem-based learning and problem-solving tools: Synthesis and direction for distributed education environments. *Journal of Interactive Learning Research, 13*(3), 239-257.

Friedman, S. L., Scholnick, E. K., & Cocking, R. R. (1987). Reflections on reflections: What planning is and how it develops. In S. L. Friedman, E. K. Scholnick, & R. R.

Cocking (Eds.), *Blueprints for thinking: The role of planning in cognitive development*, (pp. 515-534). Cambridge, MA: Cambridge University Press.

Gale, J. & Marsden, P. (1983). *Medical diagnosis: From student to clinician.* Oxford: Oxford University Press.

Ganeshan, R., Johnson, W. L., Shaw, E., & Wood, B. P. (2000). Tutoring diagnostic problem solving. In G. Goos, J. Hartmanis, & J. van Leeuwen (Eds.), *Intelligent Tutoring Systems: Fifth International Conference Proceedings,* (pp. 33-42). Berlin: Springer-Verlag.

Ge, S. & Land, S. M. (2003). Scaffolding students' problem-solving processes in an ill-structured task using question prompts and peer interactions. *Educational Technology Research and Development, 51*(1), 21-38.

Gilhooly, K. J. (1990). Cognitive psychology and medical diagnosis. *Applied Cognitive Psychology, 4*, 261-272.

Glaser, R., & Chi, M.T. (1988). Overview. In M. T. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise,* (pp. xv-xxvii). Hillsdale, NJ: Erlbaum.

Glass, G., & Hopkins, K. D. (1996). *Statistical methods in education and psychology.* Boston, MA: Allyn & Bacon.

Goldman, S. R.; Zech, L. K.; Biswas, G.; Noser, T; Bateman, H., Bransford, J.; Crews, T., Moore, A, Nathan, M., Owens, S. (1999). Computer technology and complex problem solving: Issues in the study of complex cognitive activity. *Instructional Science, 27*(3-4), 235-268.

Grant, J., & Marsden, P. (1987). The structure of memorized knowledge in students and clinicians: An explanation for diagnostic expertise. *Medical Education, 21*, 92-98.

Groen, G., & Patel, V. L. (1991). A view from medicine. In M. U. Smith (Ed.), *Toward a unified theory of problem solving: Views from the content domains,* (pp. 35-44). Hillsdale, NJ: Erlbaum.

Groen, G., & Patel, V. L. (1988). The relationship between comprehension and reasoning in medical expertise. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise*, (pp. 287-309). Hillsdale, NJ: Erlbaum.

Gruppen, L. D., Palchik, N. S., Wolf, F. M., Laing, T. J., Oh, M. S., & Davis, W. K. (1993). Medical student use of history and physical information in diagnostic reasoning. *Arthritis Care and Research, 6*(2), 64-70.

Gruppen, L. D., Woolliscroft, J. O., & Wolf, F. M. (1988). The contribution of different components of the clinical encounter in generating and eliminating diagnostic hypotheses. *Research in Medical Education: Proceedings of the 27th Annual Conference,* (pp. 242-254), November 11-17, Chicago.

Hadwin, A. (1996). Promoting self-regulation: Examining the relationships between problem-based learning and strategic content learning. *Paper Presented at the Annual Meeting of the Educational Research Association.* NY, April.

Hannan, C., Volkan, K., Fishman, M. B., Silvestri, R. C., Simon, S. R., & Fletcher, S. W. (2002). How well do second-year students learn physical diagnosis? Observational study of an objective structured clinical examination (OSCE). *BMC Medical Education, 2*(1), 1-11.

Hassebrock, F., & Prietula, M. J. *(1992).* A protocol-based coding scheme for the analysis of medical reasoning. *International Journal of Man-Machine Studies. 37*(5), 613-652.

Hativa, N., & Lesgold, A. (1991). The computer as tutor: Can it adapt to the individual learner? *Instructional Science, 20*(1), 49-78.

Hayes-Roth, B., & Hayes-Roth, F. (1979). A cognitive model of planning. *Cognitive Science, 3*, 275-310.

Henderson, E., Berlin, A., Freeman, G., & Fuller, J. (2002). Twelve tips for promoting significant event analysis to enhance reflection in undergraduate medical students. *Medical Teacher, 24*(2), 121-124.

Henderson, P., & Johnson, M. H. (2002). An innovative approach to developing the reflective skills of medical students. *BMC Medical Education, 2*(4), 1-4.

Hmelo, C. E., Gotterer, G. S., & Bransford, J. D. (1997). A theory-driven approach to assessing the cognitive effects of PBL. *Instructional Science, 25*(6), 387-408.

Hobus, P. P. M., Boshuizen, H. P. A., & Schmidt, H. G. (1990). Expert-novice differences in the mental representations of patient prototypes. *Paper Presented at the 2nd International Symposium on PBL.* Indonesia, October.

Hobus, P. P. M., Schmidt, H. G., & Boshuizen, H. P. A. (1989). Mental representation of prototypical patients: Expert –novice differences. *Paper Presented at the First European Congress of Psychology.* Amsterdam.

Hobus, P. P. M., Schmidt, H. G., Boshuizen, H. P. A., & Patel, V. L. (1987). Contextual factors in the activation of first diagnostic hypotheses: Expert-novice differences. *Medical Education, 21*, 471-476.

Hollen, A. (2000). The PBL group: Self-reflections and feedback for improved learning and growth. *Medical Teacher, 22*(5), 485-488.

Hunt, E. (1989). Cognitive science: Definition, status, and questions. *Annual Review of Psychology, 40*, 603-629.

Johnson, P. E., Duran, A. S., Hassebrock, F., Moller, J., Prietula, M., Feltovich, P. J., & Swanson, D. B. (1981). Expertise and error in diagnostic reasoning. *Cognitive Science, 5*, 235-283.

Johnson-Laird, P. (1989). *Mental models:* Cambridge, MA: Harvard University Press.

Jonassen, D. H. (1996). Scaffolding diagnostic reasoning in case-based learning environments. *Journal of Computing in Higher Education, 8*(1), 48-68.

Jonassen, D. H. & Land, S. M. (2000). *Theoretical foundations of learning environments.* Mahwah, NJ: Erlbaum.

Joseph, G. M., & Patel, V. L. (1990). Domain knowledge and hypothesis generation diagnostic reasoning. *Medical Decision Making, 10*(1), 31-46.

Kassirer, J. P. (1995). Teaching problem solving – How are we doing? *New England Journal of Medicine, 332*(22), 1507-1509.

Kassirer, J. P., & Kopelman, R. I. (1991). *Learning clinical reasoning.* Baltimore, MA: Williams & Wilkins.

Katz, S., O'Donnell, G., & Kay, H. (2000). An approach to analyzing the role and structure of reflective dialogue. *International Journal of Artificial Intelligence in Education, 11*, 320-343.

Kaufman, D. R., & Patel, V. L. (1988). Expertise in the clinical interview: A cognitive analysis. *Research in Medical Education: Proceedings of the 27th Annual Conference,* (pp. 167-172). November 11, Chicago.

Kim, S., Kolko, B., & Greer, T. H. (2002). Web-based problem solving learning: Third-year medical students' participation in end-of-life care Virtual Clinic. *Computers in Human Behaviour, 18*, 761-772.

Kintsch, W. (1992). A cognitive architecture for comprehension. In H. L. Pick, P. Van den Broek, & D. C. Knill (Eds.), *Cognition: Conceptual and methodological issues* (pp. 143-163). Washington, DC: American Psychological Association.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95,* 163-182.

Klahr, D., & MacWhinney, B. (1996). *Information processing*. Research Report (R01-DC01903): National Institute for Deafness and Communicative Disorders.

Kolodner, J. L., & Guzdial, M. (2000). Theory and practice of case-based learning. In D. H. Jonassen & S. M. Land (Eds.), *Theoretical foundations of learning environments,* (pp. 215-242). Mahwah, NJ: Erlbaum.

Koschmann, T. (2001). Revisiting the paradigms of instructional technology. *Proceedings of the Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education*, Melbourne, Australia.

Koschmann, T. (1996). Paradigm shifts and instructional technology: An introduction. In T. Koschmann (Ed.), *CSCL: Theory and practice of an emerging technology*, (pp.1-23). Mahwah, NJ: Erlbaum.

Koschmann, T., Kelson, A., Feltovich, P., & Barrows, H. S. (1996). Computer-supported problem-based learning: A principled approach to the use of computers in collaborative learning. In T. Koschmann (Ed.), *CSCL: Theory and practice of an emerging paradigm,* (pp. 83-124). Mahwah, NJ: Erlbaum.

Kosslyn, S. M. (1996). *Image and brain: The resolution of the imagery debate.* Cambridge, MA: MIT Press.

Kosslyn, S. M. & Koenig, O. (1992). *Wet mind: The new cognitive neuroscience.* NY: The Free Press.

Kotovsky, K. (2003). Problem solving--Large/small, hard/easy, conscious/non-conscious, problem-space/problem-solver: Issues of dichotomization. In J. E. Davidson & R. J. Sternberg, (Eds.), *The psychology of problem solving*, (pp. 373-383). NY: Cambridge University Press.

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior, 19*, 335-353.

Kuipers, B. & Kassirer, J. P. (1984). Causal reasoning in medicine: Analysis of a protocol. *Cognitive Science, 8,* 363-385.

Kushniruk, A. W., Patel, V.L., & Marley, A. A. J. (1998). Small worlds and medical expertise: Implications for medical cognition and knowledge. *International Journal of Medical Informatics, 49*, 255-271.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). SOAR: An architecture for general intelligence. *Artificial Intelligence, 33,* 1-64.

Laird, J. E., & Rosenbloom, P. S. (1996). The evolution of the Soar architecture. In D. Steier, & T. M. Mitchell (Eds.), *Mind matters: A tribute to Alan Newell,* (pp.1-50). Mahwah, NJ: Erlbaum.

Lajoie, S. P. (2003). Transitions and trajectories for studies of expertise. *Educational Researcher, 32*(8), 21-25.

Lajoie, S. P. (2000). *Computers as cognitive tools: No more walls*. Mahwah, NJ: Erlbaum.

Lajoie, S. P., & Azevedo, R. (2000). Cognitive tools for medical informatics. In S. P. Lajoie (Ed.), *Computers as cognitive tools: No more walls,* (pp. 247-271). Mahwah, NJ: Erlbaum.

Lajoie, S. P., Azevedo, R. & Fleiszer, D. (1998). Cognitive tools for assessment and learning in a high information flow environment. *Journal of Educational Computing Research, 18*(3) 205-235.

Lajoie, S. P., &Derry, S. (1993). *Computers as cognitive tools.* Hillsdale, NJ: Erlbaum.

Lajoie, S. P., Faremo, S., & Wiseman, J. (2001). Tutoring strategies for effective instruction in internal medicine. *International Journal of Artificial Intelligence in Education,12*(3), 293-309.

Lajoie, S. P., Lavigne, N. C., Guerrera, C., & Munsie, S. D. (2001). Constructing knowledge in the context of BioWorld. *Instructional Science, 29*(2), 155-186.

Lajoie, S. P., Lavigne, N. C., Munsie, S. D., & Wilkie, T. V. (1998). Monitoring student progress in statistics. In S. P. Lajoie (Ed.), *Reflections on statistics,* (pp.199-231). Mahwah, NJ: Erlbaum.

Lajoie, S. P., & Lesgold, A. (1989). Apprenticeship training in the workplace: Computer-coached practice environment as a new form of apprenticeship. *Machine-Mediated Learning, 3,* 7-28

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought.* NY: Basic Books.

Land, S. M. & Hannafin, M. J. (2000). Student-centered learning environments. In D. H. Jonassen, & S. M. Land (Eds.), *Theoretical foundations of learning environments,* (pp. 1-23). Mahwah, NJ: Erlbaum.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge, MA: Cambridge University Press.

Lesgold, A. (2001). The nature and method of learning by doing. *American Psychologist, 56* (11), 964-973.

Lesgold, A. (1994). Ideas about feedback and their implications for intelligent coached apprenticeship. *Machine-Mediated Learning, 4*(1), 67-80.

Lesgold, A. (1988). Problem solving. In R. J. Sternberg, & E. E. Smith (Eds.), *The psychology of human thought,* (pp. 188-213). Cambridge, MA: Cambridge University Press.

Lesgold, A., Eggan, G., Katz, S., & Rao, G. (1992). Possibilities for assessment using computer-based apprenticeship environments. In J. W. Regian & V. J. Shute (Eds.), *Cognitive approaches to automated instruction,* (pp. 49-80). Hillsdale, NJ: Erlbaum.

Lesgold, A., & Lajoie, S. (1991). Complex problem solving in electronics. In R. J. Sternberg & P. A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms,* (pp. 287-316). Hillsdale, NJ: Erlbaum.

Lesgold, A., & Nahemow, M. (2001). Learning by doing: Achieving and assessing efficient technology for learning. In S. M. Carver, & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress,* (pp. 307-346). Mahwah, NJ: Erlbaum.

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

Ley, K. & Young, D. B. (2001). Instructional principles for self-regulation. *Educational Technology Research & Development, 49*(2), 93-103.

Lillehaug, S.-I., & Lajoie, S. P. (1998). AI in medical education – another grand challenge for medical informatics. *Artificial Intelligence in Medicine, 12*, 197-225.
Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making, 14*, 331-352.
Los Arcos, J. L., Muller, W., Fuente, O., Orue, L., Arroyo, E., Leaznibarrutia, I., Santander, J. (2000). LAHYSTOTRAIN: Integration of virtual environments and ITS for surgery training. In G. Goos, J. Hartmanis, & J. van Leeuwen (Eds.), *Intelligent Tutoring Systems: Fifth International Conference Proceedings* (pp. 43-52). Berlin: Springer-Verlag.
Luger, G. (1994). *Cognitive science: The science of intelligent systems.* San Diego, CA: Academic Press.
Magnani, L. (1992). Abductive reasoning: Philosophical and educational perspectives in medicine. In D. A. Evans & V. L. Patel (Eds.), *Advanced models of cognition for medical training and practice,* (pp. 21-41). Berlin: Springer-Verlag.
Mandin, H., Jones, A., Woloschuk, W., & Harasym, P. (1997). Helping students learn to think like experts when solving clinical problems. *Academic Medicine, 72*(3), 173-179.
Mandler, J. M. (1998). Representation. In W. Damon (Ed.), *Handbook of child psychology,* (pp. 255-308). NY: Wiley.
Markman, A. B. (1999). *Knowledge representation.* Mahwah, NJ: Erlbaum.
Massaro, D., & Cowan, N. (1993). Information processing models: Microscopes of the mind. *Annual Review of Psychology, 44*, 383-425.
Mayer, R. (2003). Learning environments: The case for evidence-based practice and issue-driven research. *Educational Psychology Review, 15*(4), 359-373.
Mayer, R. (1992). Cognition and instruction: Their historic meeting with educational psychology. *Journal of Educational Psychology, 4,* 405-412.
Mayer, R. E., Dow, G. T, & Mayer, S. (2003). Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds? *Journal of Educational Psychology. 95*(4), 806-812.
McNamara, T. (1994). Knowledge representation. In R. J. Sternberg (Ed.), *Thinking and problem solving,* (pp. 81-117). San Diego, CA: Academic Press.
Miller, G. A. (1956). The magic number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.
Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior.* U.S.A.: Holt, Rinehart and Winston.
Miller, T. E. (2001). A cognitive approach to developing tools to support planning. In E. Salas & G. Klein (Eds.), *Linking expertise and naturalistic decision making,* (pp. 95-111). Mahwah, NJ: Erlbaum.
Murphy, G. C., Friedman, C. P., Elstein, A. S., Wolf, F. M., Miller, T., & Miller, J. G. (1996). The influence of a decision support system on the differential diagnosis of medical practitioners at three levels of medical training. *Proceedings of the Conference of the American Medical Informatics Association,* (pp. 219-223), October 26-30, Washington, DC.
Muzzin, L. J., Norman, G. R., Jacoby, L. L., Freightner, J. W., Tugwell, P., & Guyatt, G. H. (1983). Expertise in recall of clinical protocols in two specialty areas. *Proceedings of the 22nd Conference on Research in Medical Education,* (pp. 122-127). Washington, American Association of Medical Colleges.

Muzzin, L. J., Norman, G. R., Jacoby, L. L., Freightner, J. W., Tugwell, P., & Guyatt, G. H., (1982). Manifestations of expertise in recall of clinical protocols. *Proceedings of the 21st Annual Conference on Research in Medical Education,* (pp. 163-168). Washington, DC: American Association of Medical Colleges.

Neisser, U. (1992). Two themes in the study of cognition. In H. L. Pick, P. van den Broek, & D. C. Knill (Eds.), *Cognition: Conceptual and methodological issues.* Arlington, VA: American Psychological Association.

Neisser, U. (1967). *Cognitive psychology.* NY: Appleton-Century-Crofts.

Neufeld, V. R., Norman, G. R., Feightner, J. W., & Barrows, H. S. (1981). Clinical problem solving by medical students: A cross-sectional and longitudinal analysis. *Medical Education, 15,* 315-322.

Newell, A. (1993). The knowledge level. In P. S. Rosenbloom, J. E. Laird, & A. Newell (Eds.), *The SOAR papers: Research on integrated intelligence.* Cambridge, MA: The MIT Press.

Newell, A. (1992). Unified theories of cognition and the role of SOAR, a tribute to Alan Newell. In J. A. Michon, & A. Akyurek (Eds.), *SOAR: A cognitive architecture in perspective,* (pp. 25-79). Netherlands: Kluwer.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Newell, A., Rosenbloom, P., & Laird, J. E. (1989). Symbolic architectures for cognition. In M. Posner (Ed.), *Foundations of cognitive science,* (pp. 93-131). Cambridge, MA: MIT Press.

Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review, 65*(3), 151-166.

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

Norman, G. R., (2000). The epistemology of clinical reasoning: Perspectives from philosophy, psychology, and neuroscience. *Academic Medicine, 75*(10), S127-S133.

Norman, G. R., Brooks, L. R., & Allen, S. W. (1989). Recall by expert medical practitioners and novices as a record of processing attention. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15*(6), 1166-1174.

Norman, G. R., Brooks, L. R., Coblentz, L., & Babcook, C. J. (1992). The correlation of feature identification and category judgements in diagnostic radiology. *Memory and Cognition, 20*(4), 344-355.

Norman, G. R., Brooks, L. R., Colle, C. L, & Hatala, R. M. (1999). The benefit of diagnostic hypotheses in clinical reasoning: Experimental study of an instructional intervention for forward and backward reasoning. *Cognition & Instruction. 17*(4), 433-448.

Norman, G. R., Brooks, L. R., Smith, E. K. M., & Henry, N. L. (1987). Expert-novice differences in recall of clinical data: Resolution of a paradox. *Research in Medical Education; Proceedings of the Twenty-Sixth Annual Conference,* (pp. 69-73), Washington, DC.

Norman, G. R., & Eva, K. W. (2003). Doggie diagnosis, diagnostic success and diagnostic reasoning strategies: An alternative view. *Medical Education, 37,* 676-677.

Norman, G. R., Muzzin, L. J., Somers, S., & Rosenthal, D. (1990). Visual perception in medical practice. In Z. M. Nooman, H. G. Schmidt, & E. S. Ezzat (Eds.), *Innovation* *in medical education: An evaluation of its present status,* (pp. 204-217). NY: Springer.

Norman, G. R., Trott, A. D., Brooks, L. R., & Smith, E. K. M. (1994). Cognitive differences in clinical reasoning related to postgraduate training. *Teaching and Learning in Medicine, 6*(2), 114-120.

Orrill, C. H. (2002). Supporting online PBL: Design considerations for supporting distributed problem solving. *Distance Education, 23*(1), 41-57.

Patel, V. L. (1990). A cognitive science perspective on instructional issues. In Z. M. Nooman, H. G. Schmidt, & E. S. Ezzat (Eds.), *Innovation in medical education: An evaluation of its present status,* (pp. 189-203). NY: Springer.

Patel, V. L. & Arocha, J. F. (1995). Methods in the study of clinical reasoning. In J. Higgs & M. Jones (Eds.), *Clinical reasoning in the health professions*, (pp. 35-48). Oxford, England: Butterworth Heinneman.

Patel, V. L., Arocha, J. F., & Kaufman, D. R. (2001). A primer on aspects of cognition for medical informatics. *Journal of Medical Informatics Association, 8*, 324-343.

Patel, V. L., Arocha, J. F., & Kaufman, D. R. (1999a). Expertise and tacit knowledge in medicine. In R. J. Sternberg, & J. A. Horvath (Eds.), *Tacit knowledge in professional practice: Researcher and practitioner perspectives*, (pp. 75-99). Mahwah, NJ: Erlbaum.

Patel, V. L, Arocha, J. F, & Kaufman, D. R. (1999b). Medical cognition. In F. T. Durso (Ed.), *Handbook of applied cognition*, (pp. 663-693). Chichester, NY: Wiley.

Patel, V. L., Arocha, J. F., & Kaufman, D. R. (1994). Diagnostic reasoning and medical expertise. In D. L. Medin (Ed.), *The psychology of learning and motivation,* (pp.187-252). San Diego, CA: Academic Press.

Patel, V. L., Evans, D. A., & Groen, G. (1989). Biomedical knowledge and clinical reasoning. In D. A. Evans & V. L. Patel (Eds.), *Cognitive science in medicine: Biomedical modeling*, (pp. 49-108). Cambridge, MA: MIT Press.

Patel, V. L., Glaser, R., & Arocha, J. F. (2000). Cognition and expertise: Acquisition of medical competence. *Clinical Investigative Medicine, 23*(4), 256-260.

Patel, V. L., & Groen, G. (1991a). Cognitive frameworks for clinical reasoning: Application for training and practice. In D. A., Evans, & V. L. Patel (Eds.), *Advanced models of cognition for medical training and practice,* (pp. 193-211). Berlin: Springer-Verlag.

Patel, V. L., & Groen, G. (1991b). The general and specific nature of medical expertise: A critical look. In K. A. Ericsson & J. Smith (Eds.), *Towards a general theory of expertise: Prospects and limits,* (pp. 93-125). NY: Cambridge University Press.

Patel, V. L., & Groen, G. (1986). Knowledge based solution strategies in medical reasoning. *Cognitive Science, 10*, 91-116.

Patel, V. L., Groen, G., & Arocha, J. F. (1990). Medical expertise as a function of task difficulty. *Memory & Cognition, 18*(4), 394-406.

Patel, V. L., Groen, G., & Frederiksen, G. R. (1986). Differences between medical students and doctors in memory for clinical cases. *Medical Education, 20*, 3-9.

Patel, V. L., Groen, G., & Patel, Y. (1997). Cognitive aspects of clinical performance during patient workup: The role of medical expertise. *Advances in Health Sciences Education, 2*, 95-114.

Patel, V. L., Groen, G. J., Ramoni, M. F., & Kaufman, D. R. (1992). Machine depth versus psychological depth: A lack of equivalence. In E. Keravnou (Ed.), *Deep models for medical knowledge engineering,* (pp. 249-272). Elsevier Science.

Patel, V. L., HoPingKong, H., & Mark, B. A. (1984). Role of prior knowledge in comprehension of medical information by medical students and physicians. *Research in Medical Education: Proceedings of the 23rd Annual Conference,* (pp. 127-132), October 27 - November 1.

Patel, V. L., & Kaufman, D. R. (1998). Medical informatics and the science of cognition. *Journal of the American Medical Informatics Association, 5*(6), 493-502.

Patel, V. L., & Kaufman, D. R. (1995). Clinical reasoning and biomedical knowledge: Implications for teaching. In J. Higgs & M. Jones (Eds.), *Clinical reasoning in the health professions,* (pp. 117-128). Oxford, England: Butterworth Heinneman.

Patel, V. L., Kaufman, D. R., & Arocha, J. F. (2002). Emerging paradigms of cognition in medical decision making. *Journal of BioMedical Informatics, 35*(1), 52-75.

Patel, V. L., Kaufman, D. R., & Arocha, J. F. (2000). Conceptual change in the biomedical and health sciences domain. In R. Glaser (Ed.), *Advances in instructional psychology: Educational design and cognitive science,* (pp.329-392). Mahwah, NJ: Erlbaum.

Patel, V. L., Kaufman, D. R., & Arocha, J. F. (1995). Steering through the murky waters of a scientific conflict: Situated and symbolic models of clinical cognition. *Artificial Intelligence in Medicine, 7,* 413-438.

Patel, V. L., & Medley-Mark, B. A. (1985). Knowledge integration from clinical texts: Use of factual, inferential, and integrative questions. *Research in Medical Education: Proceedings of the 24th Annual Conference,* (pp. 91-101), October 26-31.

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

Pintrich, P. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation,* (pp. 451-502). San Diego, CA: Academic Press.

Puntembekar, S. & du Boulay, B. (1997). Design and development of MIST: A system to help students develop metacognition. *Journal of Educational Computing Research, 16*(1), 1-35.

Pylyshyn, Z. (1996). The study of cognitive architecture. In D. Steier, & T. M. Mitchell (Eds.), *Mind matters: A tribute to Allen Newell,* (pp. 51-73). Mahwah, NJ: Erlbaum.

Pylyshyn, Z. (1989). Computing in cognitive science. In M. I. Posner (Ed.), *Foundations of cognitive science,* (pp. 49-91). Cambridge, MA: MIT Press.

Ramoni, M., Stefanelli, M., Magnani, L, & Barosi, G. (1992). An epistemological framework for medical knowledge-based systems. *I.E.E.E. Transactions on Systems, Man, and Cybernetics, 22*(6), 1361-1374.

Raufauste, E., Eyrolle, H., & Marine, C. (1998). Pertinence generation in radiological diagnosis: Spreading activation and the nature of expertise. *Cognitive Science, 22*(4), 517-546.

Reimann, P., & Chi, M. T. H. (1989). Human expertise. In K. J. Gilhooly (Ed.), *Human and machine problem solving,* (pp.161-191). NY: Plenum Press.

Rikers, R. M. J. P., Schmidt, H. G., & Boshuizen, H. P. (2002). On the constraints of encapsulated knowledge: Clinical case representations by medical experts and subexperts. *Cognition and Instruction, 20* (1), 27-45.

Rikers, R. M. J. P., Schmidt, H. G., & Boshuizen, H. P. (2000). Knowledge encapsulation and the intermediate effect. *Contemporary Educational Psychology, 25*, 150-166.

Rikers, R. M. J. P., Schmidt, H. G., Boshuizen, H. P., & Linssen, G. C. M. (2002). The robustness of medical expertise: Clinical case processing by medical experts and subexperts. *The American Journal of Psychology, 115*(4), 609-630.

Rikers, R. M. J., Te Winkel, W., Loyens, S., & Schmidt, H. (2003). Clinical case processing by medical experts and subexperts. *The Journal of Psychology, 137*(3), 213-219.

Rumelhart, D. E., & Norman, D. A. (1988). Representation in memory. In R.C. Atkinson, R. J. Hernstein, G. Lindzey, & R. P. Luce (Eds.), *Steven's handbook of psychology*. Cambridge, MA: Wiley.

Saye, J. W., & Brush, T. (2002). Scaffolding critical reasoning about history and social issues in multimedia supported learning environments. *Educational Technology Research and Development, 50*(3), 77-97.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding.* Hillsdale, NJ: John Wiley & Sons.

Schmidt, H. G., & Boshuizen, H. P. A. (1993). On acquiring expertise in medicine. *Educational Psychology Review, 5*(3), 205-221.

Schmidt, H. G., & Boshuizen, H. P. A. (1991). Encapsulation of biomedical knowledge. In D. A. Evans & V. L. Patel (Eds.), *Advanced models of cognition for medical training and practice*, (pp. 265-281). Heidelberg: Springer-Verlag.

Schmidt, H. G., Boshuizen, H. P. A., & Hobus, P. P. M. (1988). Transitory stages in the development of medical expertise: The "intermediate effect" in clinical case representation studies. *Proceedings of the Cognitive Science Society Meeting.* Hillsdale, NJ: Erlbaum.

Schmidt, H. G. & Norman, G. R. (1990). Transitory stages in the development of expertise in medicine: Review of the evidence. *Paper presented at the Annual Conference of the American Educational Research Association*, Boston, MA, April 16-20.

Schmidt, H. G., Norman, G. R, & Boshuizen, H. P. A. (1990). A cognitive perspective on medical expertise: Theory and implications. *Academic Medicine, 65*(10), 611-621.

Schmidt, H. G. & Patel, V. L. (1987). Effects of prior knowledge activation through small-group discussion on the processing of science text. *Paper Presented at the Annual Meeting of the American Educational Research Association*, Washington, DC, April 20-24.

Schoenfeld, A. (1999). Looking toward the 21st century: Challenges of educational theory and practice. *Educational Researcher, 28*(7), 4-14.

Schunn, C. D., & Klahr, D. (1998). Production systems. In W. Bechtel, & G. Graham (Eds.), *A companion to cognitive science,* (pp. 542-551). Malden, MA: Blackwell.

Schuwirth, L. W. T., Blackmore, D. E., Mom, E., van den Wildenberg, F., Stoffers, H. E. J. H., & van der Vleuten, C. P. M. (1999). How to write short cases for assessing problem-solving skills. *Medical Teacher, 21*(2), 144-150.

Schwarz, S. & Griffin, T. (1986). *Medical thinking: The psychology of medical judgement and decision making.* NY: Springer-Verlag.

Shortliffe, E. H. (1991). AI meets decision support: Emerging synergies for decision support. In D. A. Evans & V. L. Patel (Eds.), *Advanced models of cognition for medical training and practice*, (pp. 71-89). Heidelberg: Springer-Verlag.

Shortliffe, E. H., Barnett, G. O., Cimino, J. J., Greenes, R. A., Huff, S. M., & Patel, V. L. (1996). Collaborative medical informatics research using the Internet and the World Wide Web. *Proceedings of the Conference of the American Medical Informatics Association,* (pp. 125-129), October 26-30, Washington D.C.

Simon, H. A. (1996). The patterned matter that is mind. In D. Steier, & T. M. Mitchell (Eds.), *Mind matters: A tribute to Alan Newell,* (pp. 407-431). Mahwah, NJ: Erlbaum.

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology, 41*, 1-19.

Simon, H. A. (1973). The structure of ill-structured problems. *Artificial Intelligence, 4*, 181-201.

Simon, H., A. & Kaplan, C. (1989). Foundations of cognitive science. In M. Posner (Ed.), *Foundations of cognitive science,* (pp. 1-48). Cambridge, MA: MIT Press.

Sisson, J. C., Donnelly, M. B., Hess, G. E., & Wooliscroft, J. O. (1991). The characterisation of early diagnostic hypotheses generated by physicians (experts) and students (novices) at one medical school. *Academic Medicine, 66*, 607-612.

Skinner, B. F. (1957). *Verbal behavior.* Englewood Cliffs, NJ: Prentice Hall.

Spivak, J. L., & Barnes, H. V. (1990). *Manual of clinical problems in internal medicine.* Boston, MA: Little, Brown & Company.

Stanovich, K. E. (2003). The fundamental computational biases of human cognition: Heuristics that (sometimes) impair decision making and problem solving. In J. E. Davidson, & R. J. Sternberg (Eds.), *The psychology of problem solving,* (pp. 291-342). NY: Cambridge University Press.

Suppes, P., Pavel, M., & Falmagne, J.-Cl. (1994). Representations and models in psychology. *Annual Review of Psychology, 45*, 517-544.

Van de Wiel, M. W. J., Boshuizen, H. P. A., & Schmidt, H. G. (1994a). Differences in knowledge application by students and medical experts in clinical reasoning. *Paper Presented at the Annual Meeting of the American Educational Research Association*, New Orleans, LA: April 4-8.

Van de Wiel, M. W. J., Boshuizen, H. P. A., & Schmidt, H. G. (1994b). The influence of task instruction and lab data format on clinical case recall. *Paper Presented at the Annual Meeting of the American Educational Research Association*, New Orleans, LA: April 4-8.

van de Wiel, M. W. J., Schaper, N. C., Boshuizen, H. P. A., & Schmidt, H. G. (1995). The explanation of encapsulating concepts by medical experts, clerks and advanced students. *Paper Presented at the Annual Meeting of the American Educational Research Association*, San Francisco, CA: April 18-22.

VanLehn, K. (1989). Problem solving and cognitive skill acquisition. In M. Posner (Ed.), *Foundations of cognitive science*, (pp. 527-579). Cambridge, MA: MIT Press.

Vincente, K. J., & Wang, J. H. (1998). An ecological theory of expertise effects in memory recall. *Psychological Review, 105*, 33-57.

Von Eckardt, B. (1993). *What is cognitive science?* Cambridge, MA: MIT Press.

Vosniadou, S. (1996). Towards a revised cognitive psychology for new advances in learning and instruction. *Learning and Instruction, 6*(2), 95-109.

Voss, J. F. & Post, T. A. (1989). On the solving of ill-structured problems. In M. T. H. Chi, R. Glaser, & M. Farr (Eds.), *The nature of expertise*, (pp. 261-285). Hillsdale, NJ: Erlbaum.

Wagner, R. (1999). Searching for determinants of performance in complex domains. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content determinants,* (pp. 371-385). Washington, D. C.: American Psychological Association.

Warner, H. R., Woolley, F. R., & Kane, R. L. (1974). Computer assisted instruction for teaching clinical decision making. *Computers and Biomedical Research, 7*, 564-574.

Wiener, N. (1948). *Cybernetics: Or, control and communication in the animal machine.* NY: Wiley.

Winn, W. (2002). Current trends in educational technology research: The study of learning environments. *Educational Psychology Review, 14*(3), 331-351.

Winne, P. H. (1997). Experimenting to bootstrap self-regulated learning. *Journal of Educational Psychology, 89*, 397-410.

Winne, P. H. & Jamieson-Noel, D. (2003). Self-regulating studying by objectives for learning: Students' reports compared to a model. *Contemporary Educational Psychology, 28*, 259-276.

Wiseman, J. (2004). Naturalistic decision making in medicine: The problem list activity. *Unpublished Master's Thesis*, McGill University, Montreal, Canada.

Wortman, P. M. (1972). Medical diagnosis: An information-processing approach. *Computers & Biomedical Research, 5*, 315-328.

Xiao, Y., Milgram, P., & Doyle, D. J. (1997). Capturing and modeling planning expertise in Anesthesiology: Results of a field study. In G. A. Klein, & C. E. Zsambok (Eds.), *Naturalistic decision making*, (pp. 197-205). Mahwah, NJ: Erlbaum.

Yip, W. (2002). Students' perceptions of the technological supports for problem-based learning. *Education and Information Technologies, 7*(4), 303-312.

Zimmerman, B. J., & Campillo, M. (2003). Motivating self-regulated problem solvers. In J. E. Davidson & R. J. Sternberg. (Eds.), *The psychology of problem solving.* (pp. 233-262). NY: Cambridge University Press.

Appendix A: Pre-Questionnaire

Participant Number:______

Examining Medical Problem Solving in a Computer-Based Learning Environment

Questionnaire: Participant Information

Please complete this brief questionnaire. All information will be kept strictly confidential.

<u>Note</u>: Please provide an estimate for any items you cannot answer exactly.

1. Age ______

2. Present level of training (tick one):
   ____ medical student, year:_____
   ____ resident, year: _____
   ____ staff physician, years on staff: _____

3. Specialization (if any) ______________________

4. Amount of classroom instruction you have received on digestive disease during your training (indicate in hours, weeks, as applicable) ______

5. Amount of experience with digestive problems in a clinic (number of hours, weeks, as applicable) ______

6. Number of digestive cases previously seen ______

7. Number of digestive cases previously diagnosed ______

8. If applicable, please describe any other experience you have had with the diagnosis of digestive disease:

____

____

____

Appendix B: Case Histories and Physical Exam Information

CASE 1: Celiac Disease

Case History

Raymond is a 27year old systems analyst from Toronto. He has experienced mild diarrhea on and off for 2 years and seen several doctors about it during that period. He has recently broken up with his girlfriend and experiences some anxiety about it. In addition to not feeling well, Raymond has lost weight in the past four months and has been feeling more tired than usual. He drinks 1 or 2 martinis per day. Raymond's last doctor told him to reduce stress in his job and to cut back on drinking coffee. Even though he has done these things the symptoms persist. Lately he has had episodes of nocturnal diarrhea that awaken him.

Physical Examination

Admitted for: dehydration, Sex: male, Age: 27

		(normal)
Temperature	37.9 C	37.9 C
Pulse	78 Bpm	60-90 Bpm
B P	110/65 MmHG	120/80 MmHG
R R	16 rpm	12-18 rpm

CASE 2: Shigellosis

Case History

Raymond, a 4 month old infant has been exhibiting fever and fussiness. His mom has given him baby aspirin to reduce the fever and thought Raymond was getting a cold. Within two days of showing the fever, Raymond began to have severe diarrhea and vomiting. His mom noticed a large amount of blood in his diapers when changing him and thought she had better start washing her hands from now on after cleaning his diapers. After 4 days, Raymond still did not seem to have a cold but was still suffering from a high fever. His mom became very worried about his condition and decided to take him to his doctor. The doctor asked his mom if there was any possibility that Raymond may have somehow gotten near his soiled diapers or put his dirty hand from touching a diaper in his mouth. His doctor recalled that he had recently prescribed antibiotics for an ear infection Raymond developed after being in contact with neighborhood kids.

Physical Examination

Admitted for: fever, Sex: male, Age: 4 months

		(normal)
Temperature	40 C	37.9 C
Pulse	155 Bpm	120-160 Bpm
B P	90/55 MmHG	100/60 MmHG
R R	46 rpm	30-50 rpm

CASE 3

Case History

Robin, having just returned from Mexico, began to feel ill and thought she would lie down. Within a few hours, Robin began to exhibit signs of nausea and vomiting. She also felt as though she had a fever. The next morning Robin awoke with some slight abdominal pain. After getting out of bed, Robin noticed she had no appetite and began to feel even more tired than before. After a few days, Robin noticed that she had discolored urine, much darker than usual, and that there was a yellowish skin discoloration around her lips and eyes.

Physical Examination

Admitted for: dehydration, Sex: female, Age: 25

		(normal)
Temperature	38.5 C	37.9 C
Pulse	100 Bpm	60-90 Bpm
B P	106/72 MmHG	120/80 MmHG
R R	18 rpm	12-18 rpm

Appendix C: Sample Case Materials

A. CASE HISTORY (3)

Directions
1. Read the following case history aloud, verbalizing your thoughts as you proceed.
2. Highlight the information in this history that you believe is relevant to a diagnosis.

NOTE: You may consult the available resources (disease library, glossary, and/or diagnostic tests) for help at any time.

CASE HISTORY (3)

Raymond is a 27year old systems analyst from Toronto. He has experienced mild diarrhea on and off for 2 years and seen several doctors about it during that period. He has recently broken up with his girlfriend and experiences some anxiety about it. In addition to not feeling well, Raymond has lost weight in the past four months and has been feeling more tired than usual. He drinks 1 or 2 martinis per day. Raymond's last doctor told him to reduce stress in his job and to cut back on drinking coffee. Even though he has done these things the symptoms persist. Lately he has had episodes of nocturnal diarrhea that awaken him.

B-1. PATIENT CHART

Case: _3______

Admitted for: ___dehydration___ Sex: __male______ Age: _27____

		(normal)	
Temp	37.9	37.9	C
pulse	78	60-90	Bpm
B P	110/65	120/80	MmHG
R R	16	12-18	rpm

B-2. DIAGNOSTIC TESTS

Directions

1. Indicate which diagnostic test(s) you wish to order from the list provided. The researcher will give you the results. You may request tests that do not appear on this form.
2. Record each test result in the appropriate cell on the form below.
3. Please provide a diagnosis (or differentials) and indicate your level of certainty (e.g. 75%).

Microbiology

TEST	CASE	NORMAL
Blood		neg
Stool		"
Stool Assay for Laxatives		"
Stool for Colistridium Difficile		"
Throat		"
Urine		"
Sputum		"
Site Discharge		"
Gram's Stain		"

Chemistry

TEST	CASE	NORMAL
Sodium		normal
Potassium		"
BUN		"
Total Bilirubin Indirect Direct		 " "
Creatinine		"
LDH		"
SGOT		"
SGPT		"
alkaline phosphotase		"
CI		"
HCO_3		"
blood sugar		"
blood ammonia		"
amelase		"

Hematology

TEST	CASE	NORMAL
Hemoglobin		normal
Hematocrit		"
Platelets		"
RBC		"
WBC-Total		"
WBC-Segs		"
WBC-Bands		"
WBC-Lymphocytes		"
WBC-Monocytes		"
WBC-Eosinophils		"
WBC-Basosinophils		"
Folic Acid		"
Mean Cell Volume		"
B_{12} level		"
Peripheral Smear		"

Serology

TEST	CASE	NORMAL
Cold Agglutinin		Neg
Hepatitis A Titer		"
Hepatitis B Titer HBsAg HBcAg HBeAg ACVAg		"
Hepatitis C Titer		"
Hepatitis D Titer		"

Serology, continued

HIV Titer		"
Influenza Titer		"
Mono Screen		"
Rubella Screen		"
Syphilis		"
T-Cell Ratios		"
pregnancy test		"
Iron		normal
Carotene		"

Urinanalysis

TEST	CASE	NORMAL
Bacteria		neg
Blood		---
Clarity		Neg
Ketones		"
Protein		"
RBC		---
WBC		---
specific gravity		normal

Other

TEST	CASE	NORMAL
% sat transforrin		"
Albumin		"
Anti-gliadin antibodies		"
Aspirin level		neg
Barium enema		"
chest X-ray		neg
colonoscopy		normal
endoscopy		"
ESR		"
forritin		"
Gastroscopy		"
physical exam		"
Plain film or 3 views of abdomen		normal
Prothrombin time		"
Serum iron		"
sigmoidoscopy		"
Small bowel biopsy		"
TIBC		normal
UGI SBFT		"
UIBC		"

C. CATEGORIZING EVIDENCE

Directions

1. Consider all of the information (evidence) you used to solve the case (your highlighted case history information, vital signs, test results, library information, etc.).
2. In the appropriate categories below, record the information that was relevant to making your diagnosis.

Patient Information

Symptoms

Diagnostic Tests

Library

Other

D. PRIORITIZING EVIDENCE

Directions

1. From C above, list the information in descending order of importance in terms of making your diagnosis (most important first).

1. ____
2. ____
3. ____
4. ____
5. ____
6. ____
7. ____
8. ____
9. ____
10. ____

E. EXPERT TRACE

Directions

1. The box below provides the ordered list of information used by an expert in diagnosing this case. Compare your ordered list of information (previous page) with the expert's.

Expert List
27 years old
chronic history
tests – koilonychia (iron deficiency)
HIV negative
preserved or increased appetite with weight loss
signs of weight loss on physical

Appendix E: Instructions to Participants

EXPERIMENTAL PROCEDURE

DIRECTIONS

During this experiment you will be asked to attempt to diagnose a set of 4 cases. For each case you will:

1. Read a case history aloud.
2. Highlight information in the case history that you believe is relevant to making a diagnosis.
3. (*optional*) Consult resources (disease library, glossary, and/or diagnostic tests).
4. Check the patient's vital signs on the patient chart.
5. Order diagnostic tests (the results will be provided by the researcher).
6. Look at all evidence you've used or generated (symptoms, test results) and categorize it (symptoms, test results, etc.)
7. Identify what information was most important to making your diagnosis.
8. Compare the information you used with what was used by an expert.

NOTE: At any point during this process you may give a diagnosis or differential diagnoses indicating your certainty level (e.g. 75%).

IMPORTANT: You are asked to think-aloud during the entire process.

Appendix F: Sample from BioWorld's On-line Library

[Shigellosis]

Description

Shigellosis is an acute intestinal infection caused by the gram-negative rod bacteria, shigella. Incubation is from 1 to 4 days. Transmission is through the fecal-oral route, by direct contact with contaminated objects or through ingestion of contaminated food or water. Shigellosis is usually accompanied by a fever in children, but not in adults. Duration can be as long as 6 weeks. Complications (from rapid dehydration and weight loss) are uncommon but can be fatal in children. Vaccines have mixed effectiveness. Prognosis is good with early proper treatment.

Symptoms

high fever in children
diarrhea
nausea
vomiting
irritability
drowsiness
abdominal pain
abdominal extension
pus, mucus, and/or blood in stools

Diagnostics

Confirmation of Shigellosis requires: a positive culture grown from stool sample and a Gram's stain test confirming gram-negative rod bacteria. Ruling out of other gram-negative bacilli, such as E. coli Microscopic bacteriologic studies also help confirm this disease. Examination of fresh stool should reveal mucus, RBC's, polymorphonuclear leukocytes. A drug sensitive test to establish effective treatment may also be needed. Other tests of body fluids and products associated with the infection may also be warranted. They can determine extensiveness of the disease or effectiveness of the treatment.

Indicators

Blood Gas:
Chemistry:
Hematology: WBC-total: high; RBC: low; Hemoglobin: high; Hematocrit: high
Microbiology: Stool: positive; Gram's Stain: gram-negative rod bacteria
Serology:
Urinalysis:

Treatment

Replacement of fluids and a low-residue diet. Antibodies are of questionable use in mild cases, but ampicillin, tetracycline, or sulfamethoxazole may be warranted in more severe cases. Anti-diarretics are contra-indicated in Shigellosis since they delay fecal excretion of the Shigellosis bacteria, thereby prolonging the fever and diarrhea.

Appendix G: Post-Questionnaire

Post-Questionnaire

Participant number ________

Directions:

Please reflect on your experience with BioWorld and complete each of the items below.

1. **In my opinion, the BioWorld cases were:**
   (1=strongly agree, 2=agree, 3=neutral, 4=disagree, 5=strongly disagree)
   1. realistic (resembled real cases) ____
   2. interesting ____
   3. difficult ____

2. **Please rate the major BioWorld activities in terms of *level of difficulty*.**
   (1=easy, 2=moderately easy, 3=moderately challenging, 4=difficult, 5=very difficult)
   ___ select symptoms from the case history
   ___ search the library
   ___ order diagnostic tests
   ___ specify a diagnosis
   ___ indicate a level of certainty
   ___ categorize evidence (putting into categories)
   ___ categorize evidence (prioritizing in order of importance)
   ___ compare your categorized evidence with that of an expert

3. Specific Comments

If you have any specific comments about any phase of BioWorld, please record them in the appropriate area:

1. problem statement: ____
   ____
2. current diagnosis/belief meter: ____
   ____
3. consult (with expert): ____
   ____
4. diagnostic tests: ____
   ____
5. library: ____
   ____
6. prioritizing evidence: ____
   ____
7. categorizing evidence (prioritizing): ____
   ____
8. comparing my evidence with that of an expert: ____
   ____

4. Other Comments

If you have any other comments regarding BioWorld or your experience using the system, please indicate what they are in the space below.

____
____
____
____

THANK YOU FOR YOUR PARTICIPATION IN THIS STUDY!