EVALUATION OF AN ONLINE ANALOGICAL PATIENT SIMULATION PROGRAM

by

Gregory Alan Thompson

A thesis submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Biomedical Informatics

The University of Utah

December 2006

THE UNIVERSITY OF UTAH GRADUATE SCHOOL

SUPERVISORY COMMITTEE APPROVAL

of a thesis submitted by

Gregory A. Thompson

This thesis has been read by each member of the following supervisory committee and by majority vote has been found to be satisfactory.

Chair: [redacted]hael J. Lincoln

Matthew H. Samore

Barry M. Stults

THE UNIVERSITY OF UTAH GRADUATE SCHOOL

FINAL READING APPROVAL

To the Graduate Council of the University of Utah:

I have read the thesis of [redacted] A. [redacted] in its final form and have found that (1) its format, citations, and bibliographic style are consistent and acceptable; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the supervisory committee and is ready for submission to The Graduate School.

[redacted]
Date

Michael J. Linc[redacted]
Chair, Supervisory Committee

Approved for the Major Department

Joyce A. Mitchell
Chair

Approved for the Graduate Council

David S. Chapman
Dean of The Graduate School

ABSTRACT

As case-based learning (CBL) via computer-assisted instruction becomes a burgeoning instructional method within medical education, its pedagogical value must be ascertained. In particular, the relative efficacy of specific instructional elements that comprise the CBL methodology must be determined. For example, numerous laboratory studies have proposed conditions that may facilitate knowledge transfer via analogy (case-based reasoning). However, few of these methods have been evaluated in complex learning environments such as medicine. The study included in this thesis, "The Medulator™ Analogical Reasoning Study" (MARS), employed an online patient simulation application to evaluate several potential methods for the optimization of learning by clinical novices (i.e., medical students).

Medulator™, a commercial Web-based patient simulation application, was modified to test the effects of case sequencing, explicit case comparison, and user-generated case summaries on user performance. Senior medical students self-enrolled via the Internet and were randomized to complete analogous sets of virtual patient cases in different sequences, with or without an explicit analogical reasoning exercise being invoked for analogous case pairs and with or without the ability to generate user-authored case summaries. Specific aspects of their case performance were then tracked. A brief follow-up user survey was conducted to determine overall satisfaction with the online CBL approach and to determine perceived value of the analogical reasoning component.

A significant effect of case sequencing on analogy transfer was seen only with respect to correct treatment scores ($p = .009$). Explicit case comparison had no reliable effect on performance. However, diagnostic accuracy increased ($p = .002$) while treatment attempts decreased ($p = .05$) when subjects were prompted to write case summaries. The additional time needed to write case summaries was not statistically significant ($p = .12$). Overall, user satisfaction with the Medulator™ was excellent. However, high perceived value of the analogical reasoning component was not supported by measured results.

Manipulating case sequences and supporting explicit case comparison yielded mixed results, suggesting that these methods afford instructional value only under specific learning conditions. However, using case summaries as a tool for reflection and proxy for self-explanation led to significant early and sustained improvement in students' performance.

This thesis is dedicated to my wife, Dr. Cheryl Thompson, without whose loving and unending support, and persuasion, this thesis work would have remained unfinished.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

The research reported in this manuscript was funded by a Small Business Innovation Research (SBIR) grant from the U.S. Department of Defense, Office of Naval Research (Contract No. N00014-03-C-0452) and awarded to Medantic Technology, LLC, a Utah Limited Liability Company, of which Dr. Thompson serves as an executive officer.

Permission to reprint the published and copyrighted article entitled "Evaluation of an Online Analogical Patient Simulation Program" was obtained from IEEE Computer Society.

CHAPTER 1

INTRODUCTION

Case-Based Learning

Case studies have traditionally been used to teach decision-making skills in professional education, as exemplified in the Harvard Business School case approach [1]. Cases are also used extensively for teaching medical science in a model called problem-based learning (PBL) whereby students work collaboratively within a mentored or facilitated small group to identify issues within the context of a problem (the case), to ask themselves germane questions, and to then identify additional information with which to answer their own questions [2]. Literature espousing the effectiveness of PBL and its core component case-based learning (CBL) is readily available in a broad array of health-care education settings [3-6]. PBL proponents argue that while the case is the stimulus in the PBL experience, actual learning occurs during the group's collaborative discussion of the case. On the other hand, PBL opponents cite the inefficiency of small group tutoring. However, CBL may also stand independently from PBL, as a subset of active (or learner-centered) learning methods, wherein learners engage with the characters and circumstances of the case. They may work in groups or independently to identify and ultimately resolve problems within realistic scenarios. As e-learning technologies infiltrate educational programs in a wide variety of disciplines, including health care, the use of the Internet to deliver CBL has grown rapidly and steadily, making independent learning through CBL highly convenient for both students and educators.

The burgeoning marriage of CBL with medical e-learning has also been associated with the use of terms such as *patient simulations* and, more recently, *virtual patient cases*. However, definitions for these terms are as broad as the e-learning technologies used to deliver CBL, from simple HTML text to virtual reality applications. For the purposes of this thesis, the use of the more common term *patient simulation* was used to generically describe high fidelity (i.e., realistic) patient cases delivered in a computer-based CBL (e-CBL) environment.

e-CBL, including CBL using high-fidelity simulation, has been established as an effective adjunctive learning method and is commonly used in medical education to compliment more traditional didactic (lecture-based) teaching methods [3–10]. Patient simulations can help students transfer knowledge from the basic sciences to patient care because they allow students to develop and practice problem-solving and decision-making skills interactively within the context of realistic patient scenarios. Another obvious benefit of e-CBL is exposing students to both uncommon diseases that they would not normally see during clinical training, as well as uncommon presentations of common diseases.

Some of the earliest studies to test the efficacy of e-CBL were conducted in the early 1990s through the University of Utah's Department of Medical Informatics employing the medical expert system, Iliad. In these studies, Iliad proved effective at teaching diagnostic skills as evidenced by students' improvement in the diagnosis of disease conditions on which they had been previously trained [7]. This effect was observed for both common and uncommon diseases on which students were previously trained [8]. In addition, it was demonstrated that generalization of the improved diagnostic accuracy occurred from specific disease conditions (e.g., Crohn's disease) to different but pathophysiologically related disease conditions (e.g., ulcerative colitis) [9]. These findings seemingly refuted the suggestion by Elstein et al. that clinical reasoning is related to domain-specific knowledge rather than a general process [11].

Practical criteria for evaluating online patient simulations have only recently been proposed [12]. Nevertheless, since the Iliad studies, several research efforts have attempted to demonstrate that e-CBL compares favorably to traditional teaching methods as measured by standard achievement tests and learner satisfaction [3–6, 10]. For example, in an uncontrolled study, Swagerty et al. showed pretest to posttest improvement in 3rd-year medical students enrolled in a case-oriented, Web-based curriculum [13]. However, media-comparative studies are also available. For example, Sakowski et al. found that students using Web-based case simulations as an individual exercise performed as well on the clerkship written examination as those in the traditional clerkship curriculum. However, this study used only a small number of subjects. Cases were not truly interactive because no user input was elicited and no feedback was given [14]. Leong et al. compared different methods of case delivery and concluded that students who used computer cases performed better on posttesting than those who studied paper cases or simple study articles [15].

Nevertheless, the overall quantity and quality of research are limited in this area of medical education. In particular, there is a paucity of studies demonstrating that the increased knowledge gained is of clinical benefit to patients. However, one recent example is provided by Fordis et al. who showed that online continuing medical education (CME), which included a case-based component, resulted in desirable changes in physician practice behavior as compared with traditional live CME [10].

However, valid arguments have been made for the past 2 decades, first by Clark [16], later by Keane et al. [17] and Friedman [18], and most recently by Cook [19], that media-comparative research is flawed due to confounding variables associated with the lack of uniformity and standardization of the media being compared (e.g., e-CBL versus classroom instruction or paper-based instruction). Instead, it is suggested that educational research should be refocused to compare like media (e.g., e-CBL versus e-CBL) in order to understand the value of specific instructional elements within these media. Others, like Fordis, reasonably counter that it is first desirable to demonstrate that one medium has *any* pedagogical benefits over traditional methods before dedicating research to that medium's specific instructional features [20].

In the spirit of this debate, this thesis is dedicated to reporting research findings from "The Medulator™ Analogical Reasoning Study" (MARS), which applied several learning principles from basic cognitive psychology research, namely, explicit case comparison, case sequencing, and case summary. These principles can be applied generically, independent of any particular instructional medium, but lend themselves especially well to e-CBL applications.

Analogical Reasoning

A critical concern in medical education is to identify and implement learning conditions that promote accurate and flexible problem solving. Particularly in problem domains involving complex dynamic situations, it is difficult or impossible to encode all the relevant knowledge in discrete rules. Work in both cognitive psychology and artificial intelligence has focused on the role of specific cases or analogies as vehicles for learning and knowledge transfer [21–24]. At the most general level, analogical reasoning (also termed *case-based reasoning*) is the process of comprehending a novel situation in terms of an already familiar situation [25]. The familiar situation (often termed the *base or source analog*) provides a type of model for making inferences about the unfamiliar situation: the target analog. The process of analogical thinking can be usefully broken into several basic constituent processes. First, when the reasoner is faced with the unfamiliar target case, one or more relevant source analogs stored in memory must be accessed and retrieved. A familiar source analog must be mapped to the target analog to identify systematic similarities between the two, thereby aligning the corresponding parts of each analog. Such similarities may include both surface (superficial or most apparent) characteristics that may be salient, although often distracting, as well as deeper structural characteristics that represent generally more relevant causal and functional relationships. The resulting mapping allows analogical inferences to be made about the target analog, thus creating new knowledge to fill gaps in understanding. These inferences need to be evaluated and possibly adapted to fit the unique requirements of the target.

Consequent to analogical reasoning, learning can result in the generation of new categories and schemas, the indexing of new instances to memory, and the new understandings of old instances and schemas that allow them to be better accessed in the future [21]. It follows then that the available fund of source analogs grows with learner experience, affording a richer case base from which to retrieve source analogs for future analogies.

It appears, however, that the process of analogical reasoning offers a number of critical steps at which inefficient mental processing can reduce the effectiveness of transfer. For example, overreliance on surface characteristics at the expense of structural characteristics may result in a poorly matched set of analogs, although competing theories of analogical reasoning have disagreed on the relative contributions of surface and structural features to the access/retrieval stage [26-27].

In 1993, Gentner, Ratterman, and Forbus published some remarkable findings from several studies related to the role of similarity among case studies in knowledge transfer, concluding that there is a "dissociation between the similarity that governs access to long-term memory and that which is used in evaluating and reasoning from a present match" [28, p. 524]. These studies build upon the universally acknowledged findings of others that analogy is central in learning and transfer. Gentner, Ratterman, and Forbus speculated on the implications of their findings for learning and transfer and described a two-staged retrieval process in a model called MAC/FAC (for "Many are called but few are chosen," from the Bible, Matthew 22). The MAC stage is a computationally cheap and fast process of selecting a number of possible analogs using a quick estimate of the similarities (both surface and structural, although surface similarities often dominate) between the target case and a group of cases in the memory pool to pass along to the FAC stage. The FAC stage is then responsible for evaluation of the best matches from the MAC stage using best structural alignment between the target and potential analogs [28].

The above-mentioned principles of reasoning may serve as an important foundation for a new pedagogical method, using case-based reasoning and e-CBL platforms as the delivery media, for improving deep memory retention and on-demand knowledge retrieval. For example, the concept of a growing case base with years of experience also suggests that the more seasoned clinician should be more proficient with analogical reasoning than the novice clinician (e.g., medical students). Indeed, the theory that novices rely more heavily on concrete causal reasoning via pathophysiologic mechanisms of disease while experts use both causal reasoning (especially for more difficult foreign analogs) as well as abstract analogical reasoning, has been proposed [29-31].

However, it remains to be determined if human novices can accelerate their acquisition of expert clinical reasoning skills through deliberate methods to teach analogical reasoning. Although most of the studies of knowledge transfer have used brief, relatively simple, narrative (i.e., noninteractive) case scenarios, in nonmedical domains (e.g., using business cases [32], college-level expository narratives [33], and even 3rd-grade science problems [34]), it is reasonable to hypothesize that analogical transfer effects also apply to cases of the sophistication and complexity levels found in medicine. Thus, the opportunity exists to apply various principles of analogy in practical e-CBL platforms and to then test the success of these new methods at effecting knowledge transfer.

About Medulator™ ("Medical Simulator")

Medulator™ (Medantic Technology, Salt Lake City, UT, USA) is a commercially available, Web-based software application, developed specifically for the authoring and delivery of medical patient simulations for teaching and assessment purposes. The target audiences for Medulator™ cases are medical students, residents, practicing physicians, and other medical professionals such as advanced practice nurses and physician assistants. The program is a data-driven, fully interactive, multimedia (graphics, medical images, and audio) Web application employing a variable response, rules-based simulation model. Friedman's *Anatomy of the Clinical Simulation* provided a useful initial framework for the functional design [35]. Working through a Medulator™ case requires the user (i.e., learner/student) to reason and function like a clinician faced with a patient who is presenting with one or more realistic medical problems. Case sessions are timed and require continuous user interaction in whatever nonprescribed sequence the user chooses, as follows:

1. In general, the user will first elicit a patient history. An intuitive Q&A format offers a selection of questions that provides patient-specific responses. To evaluate the user's effectiveness in history taking, each question can be scored for appropriateness. A physical examination is then *performed* by selecting similarly scored exam tasks. However, scored histories and physical examinations were not available features at the time of the study described in this thesis; thus, these sections were unscored.
2. The user subsequently must assimilate the patient's history and physical exam information into a list of differential diagnoses and order tests to support his or her diagnostic impressions. Test results are returned as clinical images, laboratory findings, or both.
3. The user must then select the most appropriate combination(s) of therapies to attain a positive patient outcome (highly negative outcomes are also possible). The number of therapeutic attempts is counted until a positive patient outcome is achieved.
4. A comprehensive Final Assessment section tests the user's diagnostic and treatment accuracy, followed by a multiple-format-question (e.g., multiple choice and true/false) posttest.

5. Cases may operate in either formative teaching mode or summative test mode. In formative teaching mode, interpretations of results and context-specific feedback are offered after each major case task; while in summative assessment mode, no feedback is given. In both modes, a Final Case Discussion section reviews and highlights the clinical concepts of the case that the user should have applied.

6. Finally, upon completion of each case session, a comprehensive performance review is offered to the user whereby detailed performance data are displayed for each major case section. The user's scores are compared to the averages of all users of the same case. Categories of performance indicators measured and reported include the following:

   a. Patient history score (percentage of appropriate questions selected)

   b. Physical exam score (percentage of appropriate tasks selected)

   c. Diagnostic testing (number and total costs of tests ordered)

   d. Treatment attempts (number of attempts made to achieve a positive patient outcome)

   e. Final diagnoses (correct and incorrect diagnoses selected)

   f. Final treatments (correct and incorrect definitive treatments selected)

g. Posttest score (out of five questions per case).

I designed and programmed Medulator™ using standard Web-development tools, including HTML/DHTML, JavaScript, CFML (ColdFusion MX, Macromedia Corp.), Adobe Photoshop for graphics, and MS SQL Server database.

A remote case authoring tool includes an assistive agent and clinical knowledge base that grows through the addition of each expertly authored disease-specific case. The authoring tool is integrated with a digital asset management system (DAMS) for rapidly importing preprocessed (i.e., resized and optimized for Web display) images and audio clips.

For the research described in this thesis, significant modifications were made to the standard Medulator™ case engine as follows:

1. First, an analogy transfer evaluation (ATE) was designed and added to the Final Assessment section. The ATE was designed to provide an explicit stimulated recall procedure for users to compare and contrast the current case (target analog) to all previously solved cases (source analogs). When ATE is invoked, the user must rate the similarities (and differences) between cases and justify their ratings in a simple Web form. The ATE procedure is presented to users prior to the solutions to the Final Diagnosis and Final Treatments subsections being displayed so that the current case remains an unknown target analog.

2. Second, a case summary utility was added so that users could provide for themselves a record of each case performed. Standard in Medulator™ is an automated Chart Review section that tracks and stores case data elicited and that can be accessed at any point during a case upon request through the standard navigation menu. However, the case summary utility allows users to additionally type their own impressions into a standard HTML form element. The case summary was created to serve as one of the most important methods of recall of previously solved cases when performing the ATE procedure. Thus, the case summary instructions urge the user to consider pertinent diagnostic and therapeutic concepts (structural characteristics). The case summary for each case session is stored permanently and can be retrieved by the user at anytime from the case list menu and is also provided within the ATE procedure.

Separate research of Medulator™'s pedagogical effectiveness as an e-CBL program was recently conducted in an as yet unpublished study using internal medicine residents as subjects. In this study, using a brief Medulator™-based curriculum on upper-respiratory infections (acute rhinosinusitis and acute bronchitis), the Medulator™ group performed better on test cases and posttest than residents who received traditional classroom didactic instruction covering identical material [36].

About "The Medulator™ Analogical Reasoning Study" (MARS)

MARS was conducted under U.S. Department of Defense Small Business Innovation Research (SBIR) grant funding. Research partners at the UCLA Department of Psychology and Xunesis were chosen because of their extensive prior work in cognitive psychology research, including an emphasis on analogical reasoning.

The following chapter contains the work of MARS as published in a paper entitled *Evaluation of an Online Analogical Patient Simulation Program* in the Proceedings of the Nineteenth IEEE International Symposium on Computer-Based Medical Systems. Because of the page-length restraints applied to the IEEE publication, further details of the study methodology and results are provided in Chapter 3 of this thesis, with additional discussion and conclusions provided in Chapter 4.

References

[1] Christensen, C.R., & Hansen, A.J. (Eds.). (1987). *Teaching and the case method*. Boston: Harvard Business School Press.

[2] Barrows, H.S., & Tamblyn, R.M. (1980). *Problem-based learning: An approach to medical education*. New York: Springer.

[3] Thomas, M.D., O'Connor, F.W., Albert, M.L., Boutain, D., & Brandt P.A. (2001). Case-based teaching and learning experiences. *Issues Ment Health Nurs, 22*(5), 517-531.

[4] Schwartz, R.W., Donnelly, M.B., Nash, P.P., Johnson, S.B., Young, B., & Griffen, W.O., Jr. (1992). Problem-based learning: An effective educational method for a surgery clerkship. *J Surg Res, 53*(4), 326-330.

[5] Neistadt, M.E., Wight, J., & Mulligan, S.E. (1998). Clinical reasoning case studies as teaching tools. *Amer J Occup Ther, 52*(2), 125-132.

[6] Enarson, C., & Cariaga-Lo, L. (2001). Influence of curriculum type on student performance in the United States Medical Licensing Examination Step 1 and Step 2 exams: Problem-based learning vs. lecture-based curriculum. *Med Educ, 35*(11), 1050-1055.

[7] Turner, C.W., Lincoln, M.J., Haug, P., Williamson, J.W., Jessen, S., Cundick, K., & Warner, H. (1991). Iliad training effects: A cognitive model and empirical findings. *Proc Annu Symp Comput Appl Med Care,* 68-72.

[8] Lincoln, M.J., Turner, C.W., Haug, P.J., Warner, H.R., Williamson, J.W., Bouhaddou, O., Jessen, S.G., Sorenson, D., Cundick, R.C., & Grant, M. (1991). Iliad training enhances medical students' diagnostic skills. *J Med Syst, 15*(1), 93-110.

[9] Lincoln, M.J., Turner, C.W., Haug, P.J., Williamson, J.W., Jessen, S., Cundick, R.M., Cundick, K., & Warner, H.R. (1992). Iliad's role in the generalization of learning across a medical domain. *Proc Annu Symp Comput Appl Med Care,* 174-178.

[10] Fordis, M., King, J.E., Ballantyne, C.M., Jones, P.H., Schneider, K.H., Spann, S.J., Greenberg, S.B., & Greisinger, A.J. (2005). Comparison of the instructional efficacy of Internet-based CME with live interactive CME workshops: A randomized controlled trial. *JAMA, 294*(9), 1043-1051.

[11] Elstein, A.S., Shulman, L.S., & Sprafka, S.A. (1990). Medical problem solving, a ten year retrospective. *Eval and the Health Professions, 13,* 5-36.

[12] Shim, B., Brock, D., & Jenkins, L. (2005). Developing practical criteria for evaluating online patient simulations: A preliminary study. *Med Teach, 27*(2), 175-177.

[13] Swagerty, D., Jr., Studenski, S., Laird, R., & Rigler, S. (2000). A case-oriented Web-based curriculum in geriatrics for third-year medical students. *J Amer Geriatr Soc, 48*(11), 1507-1512.

[14] Sakowski, H.A., Rich, E.C., & Turner, P.D. (2001). Web-based case simulations for a primary care clerkship. *Acad Med, 76,* 547.

[15] Leong, S.L., Baldwin, C.D., & Adelman, A.M. (2003). Integrating Web-based computer cases into a required clerkship: Development and evaluation. *Acad Med, 78*(3), 295-301.

[16] Clark, R. (1985). Confounding in educational computing research. *J Educ Comput Res, 1,* 28–42.

[17] Keane, D., Norman, G., & Vickers, J. (1991). The inadequacy of recent research on computer-assisted instruction. *Acad Med, 66,* 444–448.

[18] Friedman, C. (1994). The research we should be doing. *Acad Med, 69,* 455–457.

[19] Cook, D.A. (2005). The research we still are not doing: An agenda for the study of computer-based learning. *Acad Med, 80*(6), 541-548.

[20] Fordis, M. (2006). Reply to letter to the editor by D.A. Cook. *JAMA, 295*(7), 758-759.

[21] Gentner, D., & Holyoak, K.J. (1997). Reasoning and learning by analogy: Introduction. *Amer Psychol, 52,* 32-34.

[22] Holyoak, K.J., & Thagard, P. (1995). *Mental leaps: Analogy in creative thought.* Cambridge, MA: MIT Press.

[23] Kolodner, J.L. (1997). Educational implications of analogy. A view from case-based reasoning. *Amer Psychol, 52*(1), 57-66.

[24] Gentner, D., Holyoak, K.J., & Kokinov, B.N. (Eds.). (2001). *The analogical mind: Perspectives from cognitive science.* Cambridge, MA: MIT Press.

[25] Gentner, D. (1983). Structure mapping: A theoretical framework for analogy. *Cognitive Science, 7*(2), 155-170.

[26] Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development, 67,* 2797-2822.

[27] Goldstone, R.L., & Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cognitive Psychol, 46,* 414-466.

[28] Gentner, D., Rattermann, M., & Forbus, K. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychol, 25,* 4524-4575.

[29] Schmidt, H.G., Norman, G.R., & Boshuizen, H.P.A. (1990). A cognitive perspective on medical expertise: Theory and implications. *Acad Med, 65,* 611-621.

[30] Patel, V.L., Arocha, J.F., & Zhang, J. (2005). Thinking and reasoning in medicine. In K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning.* New York: Cambridge University Press.

[31] Coderre, S., Mandin, H., Harasym, P.H., & Fick, G.H. (2003). Diagnostic reasoning strategies and diagnostic success. *Med Educ, 37,* 695–703.

[32] Gentner, D., Lowenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *J Educ Psychol, 95,* 393-408.

[33] Yanowitz, K.L. (2001). The effects of similarity of theme and instantiation in analogical reasoning. *Amer J Psychol, 114*(4), 547-567.

[34] Yanowitz, K.L. (2001). Using analogies to improve elementary school students' inferential reasoning about scientific concepts. *School Science and Mathematics, 101*(3), 133-142.

[35] Friedman, C.P. (1995). Anatomy of the clinical simulation. *Acad Med, 70*(3), 205-209.

[36] Milne, C., Benson, S., Samore, M., Stoddard, G., Stults, B., & Thompson, G.A. (2006). *The use of Web-based computer-based learning to replace required conferences in IM training programs.* Unpublished manuscript, University of Utah, Salt Lake City.

CHAPTER 2

EVALUATION OF AN ONLINE ANALOGICAL PATIENT SIMULATION PROGRAM

Authors

Gregory A. Thompson, MD, University of Utah, Department of Medical Informatics and Medantic Technology, LLC, Salt Lake City, UT

Robert G. Morrison, PhD, Xunesis

Keith J. Holyoak, PhD, UCLA Department of Psychology

Terry K. Clark, MD, Medantic Technology, LLC, Salt Lake City, UT

Grant Support

Small Business Innovation Research grant awarded 10.01.2003 to Medantic Technology, LLC, by the U.S. Department of Defense, Office of the Secretary of Defense, Office of Naval Research (Contract No. N00014-03-C-0452)

Published

Thompson, G.A., Morrison, R.G., Holyoak, K.J., & Clark, T.K. (2006). Evaluation of an online analogical patient simulation program. *Proceedings of the Nineteenth IEEE International Symposium on Computer-Based Medical Systems* (pp. 623-628). Los Alamitos, CA: IEEE Computer Society.

Abstract

Medulator™, a commercial Web-based, variable response, patient simulation application, was modified to test the effect of case sequencing, explicit case comparison, and user-generated case summaries on overall user performance. METHODS: Senior medical students completed analogous sets of virtual patient cases in different sequences, and their case performance was tracked. A follow-up user Satisfaction Survey was conducted. RESULTS: A significant effect of case sequencing on analogy transfer was seen only with respect to correct treatment scores ($p = .009$). Explicit case comparison had no reliable effect on performance. However, diagnostic accuracy increased ($p = .002$) while treatment attempts decreased ($p = .05$) when subjects were prompted to write case summaries. Satisfaction with the patient simulation program was high. CONCLUSION: Manipulating case sequences and supporting explicit case comparison yielded mixed results. However, using case summaries as a tool for reflection and proxy for self-explanation led to significant improvement in students' performance.

Introduction

Analogical (case-based) reasoning is ubiquitous in real-world medical diagnosis. Yet most technology-enhanced medical training systems fail to provide new information and training in a manner consistent with the way professionals need to later access learned information. Medulator™ (Medantic Technology, Salt Lake City, UT, USA) provides an ecologically valid alternative to these systems and also provides the opportunity to optimize learning through application of principles of transfer from the analogical reasoning literature. Numerous laboratory studies have proposed conditions that may facilitate transfer via analogy [see 1–3 for reviews]; however, few of these methods have been evaluated in complex learning environments such as the domain of medicine. In the present study, we use Medulator™ to evaluate two potential methods for the optimization of learning by novices (i.e., students).

Analogical reasoning involves comparison of structured information (i.e., the pattern of relationships) between two cases and allows the reasoner to make inferences about one case (the *target*) based on prior knowledge of another case (the *source*). For example, the diagnoses for two patients may be said to be analogous if they have similar patterns of symptoms, physical findings, and diagnostic test results. However, the objects in the source and target of an analogy can also be similar at a surface level (e.g., two patients may be the same gender, race, age, and have similar presenting complaints). These nondiagnostic surface characteristics can frequently be quite salient and can distract reasoners from a full appreciation of the structural similarities between two cases. Thus, analogy transfer to novel cases will be promoted if the learner is led to focus on structural similarities.

There is some debate in the experimental literature as to whether surface similarities that correlate with structural similarities may aid in initial learning. On one hand, the salience of surface similarities may facilitate initial detection of the less salient structural similarities, at least for young children [4]; however, the presence of these surface similarities may in some circumstances lead the learner to overlook the diagnostic structural characteristics [5]. In this study, we investigate these alternatives by varying the order of cases with respect to surface and structural similarity.

A second factor that has been shown to affect analogical transfer and learning under certain circumstances is explicit case comparison during study [6–8]. For example, Gentner, Lowenstein, and Thompson [8] demonstrated that business students were more likely to recall analogically relevant source cases when they were required to explicitly compare cases during study. However, the effectiveness of this strategy may be domain specific, both because of the way knowledge in a domain is structured and because different types of learners may implicitly use analogy as a standard learning mechanism. For instance, medical and legal professionals who work in a domain that is dominated by case-based reasoning may be less sensitive to explicit comparison enhancement than business students who work in a domain that is not as structured with respect to cases. To evaluate the effect of explicit comparison during study, we modified Medulator™'s Final Assessment section to include an explicit analogy transfer evaluation (ATE). The ATE requires learners to compare and contrast the current case to previous known cases (at least one of which is a true structural analog of the current case).

An integral part of ATE is the case summary component, which students use as a self-reminder of previously solved cases' germane features when comparing and contrasting to an unknown case. Literature also shows that using self-explanation in problem-solving tasks improves performance [9–12]. In the context of ATE, case summary serves as a self-explanation proxy. Thus, a separate arm of this study examined the effect of learner-generated case summaries on learner performance, independent of ATE.

Research Objectives

1. Determine whether case ordering that manipulates the relative surface and structural similarity between adjacent cases affects learning as measured by Medulator™ performance metrics.
2. Determine whether explicit comparison as implemented in the ATE can enhance learning as measured by Medulator™ performance metrics.
3. Determine effect of case summaries on learning as measured by Medulator™ performance metrics.
4. Determine user satisfaction with Medulator™ and perceived effect of ATE on the diagnostic process.

Methods

We used Medulator™ to study the effect of case sequencing, explicit case comparison, and writing case summaries on diagnostic and treatment performance of clinical novices using cases that systematically varied structure (i.e., diagnosis determinants such as full symptom constellation, physical examination findings, diagnostic test results, and response to therapy) and surface characteristics (i.e., salient, nondiagnostic information such as patient age, gender, occupation, chief complaint, and presenting symptoms). Subjects were senior medical students.

Each subject worked through 11 physician-authored Medulator™ virtual patient cases. Diagnoses were from one of three groups: (1) four analogous bioterrorism cases of primary lower respiratory infections (anthrax, pneumonic plague, Q fever, and tularemic pneumonia); (2) four analogous cardiology cases of congestive heart failure (CHF) (hypertensive CHF, idiopathic dilated cardiomyopathy with CHF, acute myocardial infarction with CHF, and infective endocarditis with CHF); and (3) three nonanalogous distracter cases. Cases were structurally analogous *within* their own diagnostic category but superficially similar *within and/or across* diagnostic categories.

The same 11 cases were presented in one of two different orderings. In the "easy" ordering, cases that had similar structural and surface characteristics were presented earlier in the sequence; while in the "hard" ordering, early cases shared only structural characteristics and not surface characteristics. As a result, both groups of subjects saw identical cases and, most importantly, the test cases (#9 and #11) were identical between groups.

Ninety-six (96) senior medical students who had never used Medulator™ self-enrolled over the Web and were paid $150 for their participation, which took 5.0 hours on average. Subjects were randomly assigned to one of six groups (see Table 1).

Subjects completed cases in a defined sequence (easy or hard). For Groups 1 and 4, explicit case comparison was invoked via ATE on the 2nd, 3rd, and 4th (test case) analogs of each analogous case set. In those two groups, students were

Table 1. Randomized study group assignment

Group	Case Ordering	ATE	Case Summary
1	Hard	Yes	Yes
2	Hard	No	Yes
3	Hard	No	No
4	Easy	Yes	Yes
5	Easy	No	Yes
6	Easy	No	No

encouraged to write a case summary for every case to serve as a reminder of the cases' salient features. All other groups received no instructions for comparing previous cases but proceeded directly to the Final Assessment section (final diagnosis and treatment selections). However, Groups 2 and 5 also wrote case summaries (without ATE) for each case and Groups 3 and 6 did not write case summaries. Subjects wrote case summaries by typing into a free-text space and submitting for later retrieval.

The measured dependent variables were:

1. Number of treatment attempts required to achieve a positive patient outcome
2. Number of correct/incorrect diagnoses chosen
3. Number of correct/incorrect treatments chosen
4. Number and total costs of diagnostic tests chosen
5. Case time (keyboard time).

Subjects were initially allowed 2 weeks to complete the study. However, in order to achieve the goal of at least 64 completions, some subjects were granted up to three 1-week extensions. An honor system was published stating that subjects would work independently and with no external assistance.

ATE Methodology

The ATE condition consisted of the following:

1. First, once students completed selecting their final diagnoses and final treatments, they were presented with instructions to rate the degree to which previously completed cases were structurally similar to the current case. Upon submitting their ratings, students were given feedback as to which case(s) were the closest analogs (as determined by the case authors).
2. Next, students were asked to select the categories in which the case analogs were most similar, then most different. Eight structural categories were offered for comparison and contrast, including symptom constellation, pertinent diagnostic tests, effective treatments, and so on. Students were then asked to justify their responses in free text. Upon submission of this page, an expert opinion of the analogies was given, which students were expected to use to mentally index the current analog for future reference.
3. Finally, the correct diagnoses and treatments for the case were revealed with feedback on the student's selections.

Satisfaction Survey Methods

Following completion of the 11 cases, subjects were asked to complete a simple user Satisfaction Survey online. All subjects were asked to respond to four core questions with Likert scale responses. ATE subjects were asked to respond to an additional two questions specifically related to their ATE experience. Space was also provided for free-text comments.

Results

Of the 96 subjects who originally self-enrolled, 72 subjects completed all 11 cases (33 in ATE group, 39 controls). Only data from subjects who completed all 11 cases were analyzed. Outlying data were discarded using a three standard deviation cut.

Case Ordering

To measure the effect of case order (easy versus hard) on performance, we analyzed test case #9 (a bioterrorism case) and case #11 (a cardiology case). To control for the effect of case summaries, Groups 3 and 6 were eliminated from analysis; thus, all subjects in this analysis wrote case summaries for every case. Half of the subjects in this analysis compared cases using the ATE instructions (Groups 1 and 4) while half performed just case summaries (Groups 2 and 5). We performed a 2 x 2 between-subject analysis of variance (ANOVA) to investigate the effect of case sequencing and explicit comparison (ATE), as well as their possible interaction. Although main effects of case order and ATE were not statistically significant, a reliable crossover interaction between case order and ATE was seen ($p = .009$) on the correct treatment score (see Figure 1). If case type (bioterrorism versus cardiology) is included in the analysis, there is a trend ($p = .13$) towards a 3-way interaction, which appears to be driven more by the cardiology case than by the bioterrorism case.

Explicit Case Comparison (ATE)

To evaluate the effectiveness of explicit comparison between cases as implemented using the ATE, we included only cases in which the ATE group subjects were given ATE instructions; thus, the first case in each diagnostic category as well as the distracter cases were excluded. Groups who did not write case summaries (Groups 3 and 6) were not included in this analysis. We performed a mixed 2 x 2 ANOVA that factored diagnostic category (i.e., cardiology versus

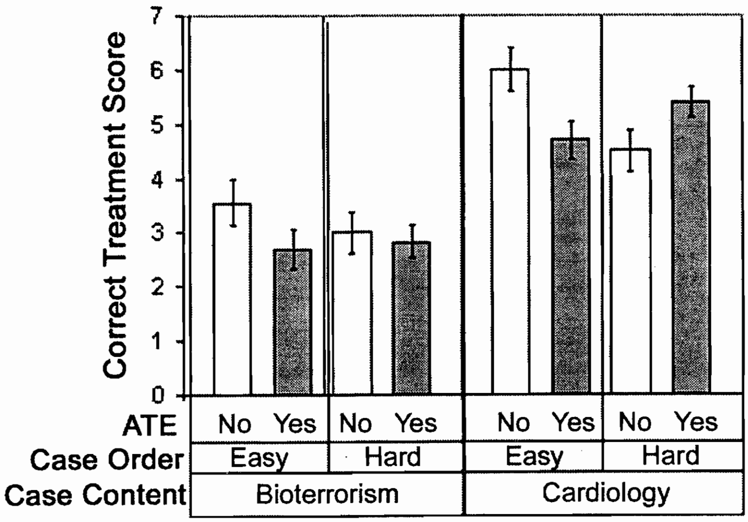


Figure 1. Three-way interaction of case type with case ordering and ATE procedure. Error bars represent ± 1 standard error.

bioterrorism cases) as a within-subject factor and explicit comparison (ATE versus no ATE) as a between-subject variable. Subjects spent more on tests (i.e., test costs) on bioterrorism cases ($p = .03$), and there was a trend ($p = .05$) towards a reduction in test costs in the ATE condition. Likewise, diagnosis and treatment were harder with bioterrorism cases ($p < .001$), but there was no reliable effect of ATE. Treatment attempts (i.e., number of measured treatment attempts users required to achieve a positive patient outcome) were higher with ATE ($p = .013$). ATE produced no significant effect on study and solution time (case time).

Case Summaries

To measure the effect of case summary on performance, independent of ATE, we looked at all cardiology and bioterrorism cases. Groups 2 and 4 were eliminated from analysis because they did ATE in addition to case summaries. We performed a mixed 2 x 2 ANOVA that factored diagnostic category (i.e., cardiology versus bioterrorism cases) as a within-subject factor and whether or not subjects were required to write case summaries as a between-subject variable. Diagnostic accuracy reliably increased ($p \leq .002$) while treatment attempts decreased ($p = .05$) when subjects were prompted to write case summaries. A trend suggested that writing case summaries improved diagnosis of bioterrorism cases more than cardiology cases ($p = .06$) and treatment attempts mainly decreased with case summary writing for cardiology cases ($p = .04$). Writing case summaries had no reliable effect on test costs or treatment score.

User Satisfaction Survey Results

Seventy-one (71) subjects completed the Satisfaction Survey, 33 of which were ATE subjects. Sixty-five percent (65%) of subjects responded that they had used similar patient simulation tools only rarely (once per year or less) or never. Overall, 93% rated Medulator™ 4 (very good) or 5 (excellent), and 73.2% rated 4 (highly) or 5 (completely) for applicability of Medulator™ to their training. User comments indicated that applicability was reduced due to the concentration of bioterrorism cases rather than more "common" medical cases.

Ninety-seven percent (97%) of ATE subjects responded that Medulator™ was moderately or very effective at helping them to think analogically, and 75.8% felt that using analogical reasoning was moderately or very effective in helping them to solve cases.

Discussion

In this study, we attempted to apply several principles from the analogy basic research literature to enhance medical learning using Medulator™, an online, interactive, multimedia patient simulator. Specifically, we investigated whether using explicit comparison of cases through ATE would increase students' ability to identify relevant diagnostic and therapeutic principles, improving their clinical accuracy or efficiency. We also investigated case ordering, hypothesizing that forcing students to focus on structural characteristics without the support of nondiagnostic surface characteristics might ultimately immunize them from distraction by surface characteristics and improve their performance. Last, we evaluated whether writing case summaries would improve performance on analogous cases.

Results from this study were mixed. Although ATE may improve cost effectiveness of diagnostic work-up (i.e., lower test costs), there was not strong evidence that applying explicit comparison of cases using ATE had a positive influence on measures of diagnostic performance. Likewise, case ordering did not have a reliable effect on test-case performance. However, the positive interaction seen between case ordering and ATE with respect to correct treatment score may suggest that explicit comparison can improve performance when students encounter difficult cases first. One explanation for this result is that explicit comparison of cases tends to focus students on all salient characteristics of the case—both diagnostic structural characteristics and nondiagnostic surface characteristics. In the "easy" case ordering, surface and structural characteristics were aligned across early cases; thus, students may have mistakenly associated surface characteristics with diagnostic efficacy. In contrast, when nondiagnostic surface characteristics do not align with diagnostic structural characteristics in early cases, explicit comparison via the ATE seems to improve later treatment performance. This appears to be particularly true for cases in which the students may already be familiar with the diagnostic domain (i.e., cardiology rather than bioterrorism cases). Specifically, for cases in which the students are more familiar with the treatment principles, ATE instructions focus them on structural comparisons that the "hard" case ordering encourages. This suggests that for new areas of medical learning it is necessary to first educate students on important treatment principles before moving on to case-based learning (CBL) methods.

One reason why this interaction was seen on correct treatment scores, and not correct diagnosis scores, may be that medical students tend to rely more heavily on causal reasoning when generating differential diagnoses. It is possible that medical diagnosis for medical students may not be principally analogical in nature but rather driven by pathophysiological correlations [13]. However, once a diagnosis is correctly determined, relevant exemplar cases may be useful for determining treatment analogically. This explanation would then make it unlikely that ATE failed to improve subjects' diagnostic performance because medical students (in contrast to Gentner, Lowenstein, and Thompson's business students [8]) instinctively think analogically; therefore, forcing explicit comparison of cases is superfluous. It is more likely that the students participating in this study did not have the necessary expertise to abstract the diagnostic principles from the cases. Thus, explicit comparison of cases did not reinforce these principles. (See Chi, Feltovich, and Glaser [14] for a similar issue in the domain of physics.) In contrast, expert clinicians may rely on a repertoire of cases, or "illness scripts," built from personal clinical experience when applying diagnostic reasoning [15].

There are other potential reasons why ATE did not have a more significant effect on overall case performance. It may be that subjects' memory of analogs simply decayed over time: A per-subject analysis of case performance against time between study initiation and study completion would be complicated and has not been attempted. Although this explanation is possible, medical professionals obviously use very old knowledge from previous cases to diagnose new cases. Thus, this explanation may again interact with the experience level of the student/professional.

Last, explicit comparison may have failed to produce a greater effect because of the complexity and interactive nature of the case analogs themselves. Previous studies on the effect of analogical reasoning have used relatively simplistic case scenarios with fewer variables to consider and categorize as surface versus structural characteristics [e.g., 8]. Also, in those studies, information was passively transferred, such that subjects were assured of being exposed to all structural characteristics germane to solving the problem. In contrast, cases that offer comprehensive detail such as the medical cases used in this study may present too many variables to make definite determinations about the similarities and differences among cases. Subjects could be overwhelmed when trying to determine which structural categories of data (e.g., historical, diagnostic, and therapeutic) are most important to compare and contrast. Furthermore, due to the highly interactive nature of Medulator™ cases (emulating real-world information gathering), subjects may fail to elicit certain critical structural information required for source analog comparison.

There were some notable limitations to the present study. Human factors could have led to overestimation of some dependent variables in all groups. For example, when selecting diagnostic tests, correct diagnoses, or correct treatments, subjects could potentially take a "shotgun approach" by selecting more options than necessary in order to observe the feedback or to ensure a correct selection, leading to spuriously high numbers of incorrect selections.

Another potential limitation was the variability in the time period over which subjects completed cases. Subjects were initially urged to complete all cases within 2 weeks. Previous laboratory studies requiring analogical reminding have typically been conducted over 1 week or less [9]. Justification for this stipulation is founded in the concept that analogical reasoning requires the subject to be able to recall the pertinent structures of known cases. Case summaries served this purpose to the extent that subjects were insightful and diligent in their self-generated accounts of each case. Nevertheless, when excessive periods of time transpire between the source case and the target case, an unpredictable degree of information decay can occur, limiting the usefulness of the subject's recall. However, in order to reach the target of subject completions (and avoid participant dropout), study account extensions were granted in 1-week increments. Thus, between-subject variance may have dramatically increased because some students completed all cases within a brief period of time (e.g., 24 hours) versus others who took more than 1 month to complete all 11 cases, resulting in weeks between their first and last cases.

Of note, and of particular interest to designers of patient simulation programs, is case performance improvement resulting from the use of case summaries as a tool for reflection and as a proxy for self-explanation. In the present study, diagnostic accuracy improved and treatment attempts were reduced in subjects who summarized in writing the pertinent information in each case. This effect was observed early (from the first case in each sequence) and was sustained throughout each case sequence. Although this result is consistent with previous research [1-4], and in itself is not surprising—after all, case summaries are analogous to physicians writing an "H&P" or a "SOAP note"—these investigators are aware of no comparable computer-based patient simulation platforms that incorporate such a simple yet effective tool. Furthermore, informal analysis of subjects' case summaries written for this study revealed that even subjects in groups who did not perform ATE (and, therefore, did not need case summaries as a recall tool) wrote word counts comparable to ATE subjects, suggesting that those subjects were using the case summaries for reflection and self-explanation.

Finally, the follow-up survey results demonstrate a high degree of user acceptance of Medulator™. Subject experience with patient simulation tools such as Medulator™ is largely lacking, with 65% of subjects responding that they have used similar tools only rarely (once per year or less) or never. Although the cases in this study operated in pure assessment mode, user comments indicated that many considered Medulator™ an excellent learning tool—an observation supported by their improvement in performance during the course of study. However, subjects perceived explicit analogical reasoning to be more effective than it was. Nearly all (97%) of the ATE subjects felt that Medulator™ was at least somewhat effective at helping them to understand the concept of analogical (case-based) reasoning.

Three-fourths of ATE subjects also felt that analogical reasoning was at least somewhat helpful to them in solving cases, even though objective measurements did not agree. (ATE was not statistically associated with improved diagnostic accuracy.) Interestingly, the applicability of bioterrorism cases to subjects' training was rated low.

In conclusion, we found that the modifications made to Medulator™ for manipulating case sequences and supporting explicit case comparisons yielded mixed results. There may be evidence that such conditions have a greater effect on therapeutic reasoning than diagnostic reasoning, primarily when applied in familiar knowledge domains. However, using case summaries as a tool for reflection and proxy for self-explanation led to significant improvement in students' performance.

References

[1] Holyoak, K.J., & Thagard, P. (1995). *Mental leaps: Analogy in creative thought.* Cambridge, MA: MIT Press.

[2] Gentner, D., Holyoak, K.J., & Kokinov, B.N. (Eds.). (2001). *The analogical mind: Perspectives from cognitive science.* Cambridge, MA: MIT Press.

[3] Holyoak, K.J. (2005). Analogy. In K.J. Holyoak & R.G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 117-142). Cambridge, UK: Cambridge University Press.

[4] Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development, 67,* 2797-2822.

[5] Goldstone, R.L., & Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cognitive Psychol, 46,* 414-466.

[6] Gick, M.L., & Holyoak, K.J. (1983). Schema induction and analogical transfer. *Cognitive Psychol, 15,* 1-38.

[7] Catrambone, R., & Holyoak, K.J. (1989). Overcoming contextual limitations on problem-solving transfer. *J Exper Psychol: Learning, Memory, and Cognition, 15,* 1147-1156.

[8] Gentner, D., Lowenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *J Educ Psychol, 95,* 393-408.

[9] Bielaczyc, K., Pirolli, P.L., & Brown, A.L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction, 13*(2), 221-252.

[10] Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 5,* 145-182.

[11] Neuman, Y., & Schwarz, B. (1998). Is self-explanation while solving problems helpful? The case of analogical problem solving. *British J Educ Psychol, 68,* 15-24.

[12] Chi, M.T.H., Deleeuw, N., Chiu, M.H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 182*(3), 439-477.

[13] Patel, V.L., Arocha, J.F., & Zhang, J. (2005). Thinking and reasoning in medicine. In K.J. Holyoak & R.G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning.* New York: Cambridge University Press.

[14] Chi, M.T.H., Feltovich, P.J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121-152.

[15] Sisson, J., Donnelly, M., Hess, G., & Woolliscroft, J. (1991). The characteristics of early diagnostic hypotheses generated by physicians (experts) and students (novices) at one medical school. *Acad Med, 66,* 607-612.

CHAPTER 3

ADDITIONAL MARS METHODS AND RESULTS

Due to space restrictions applied to the IEEE publication, this chapter elaborates on the publication by providing additional Methods and Results. Additional Discussion and Final Conclusions are provided in Chapter 4.

Additional MARS Methods

Analogous Diagnostic Categories

It seems warranted to explain the choice of diagnoses for the analogous cases used in MARS. It was desirable to choose diagnostic categories that are potentially confusable to the clinical reasoner so that the reasoner would be motivated to explore the structural characteristics of the cases in order to make a specific diagnosis. On the surface, respiratory infections and congestive heart failure (CHF) can have similar patient presentations (e.g., similar chief complaints; similar time course; and similar symptoms of cough, dyspnea, and fatigue). Thus, if the clinical reasoner were to rely upon surface characteristics as deterministic of diagnosis, a case of pneumonia could be easily confused with a case of hypertensive CHF, for example. It would only be upon further investigation into structural characteristics that a more specific diagnosis would become apparent with additional patient history (e.g., pleuritic chest pain in pneumonia versus orthopnea and peripheral edema in CHF); physical examination (e.g., fever and localized rhonchi in pneumonia versus diffuse rales, heart murmur, and pitting edema in CHF); diagnostic testing (e.g., consolidation on chest x-ray in pneumonia versus pulmonary edema with or without cardiomegaly in CHF); and even empiric therapy (e.g., antibiotics for pneumonia versus diuretics and ACE inhibitors for CHF).

The reason that bioterrorism cases were used for the upper respiratory infection analogs is simple: The study was funded by the U.S. Department of Defense (DoD) and bioterrorism is a knowledge domain of great interest to DoD grant managers. Nevertheless, the bioterrorism component added an important level of difficulty that made it possible to distinguish effects of the independent variables on performance with respect to knowledge domain familiarity. In other words, senior medical students generally have a good foundation of knowledge for most common causes of CHF but not for bioterrorism. However, the bioterrorism component was essentially an epidemiologic surface feature, since the structural features of those cases dealt with the primary respiratory infection aspect, and senior medical students should be familiar with the diagnosis and management of many types of pulmonary infection. Thus, it was not believed that the bioterrorism aspect added a significant confounding factor.

Case-Ordering Method

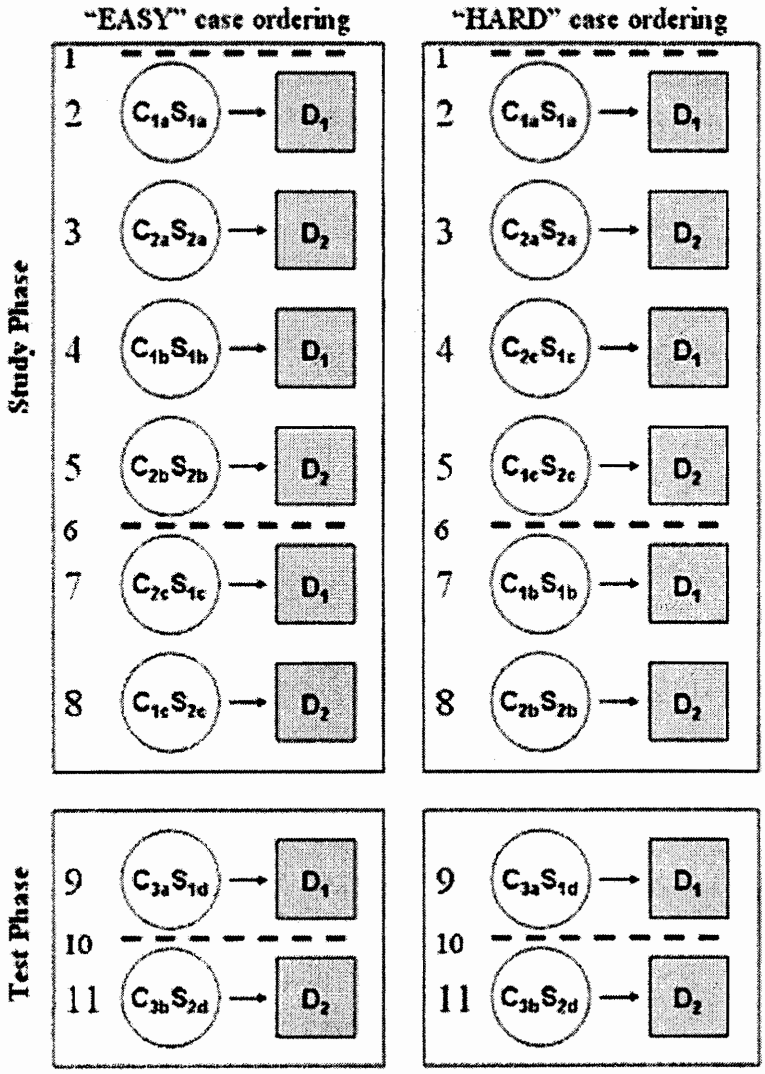
Next, a more explicit explanation of the case ordering used may be helpful. To reiterate, the case-ordering manipulation was used to measure the relative effects of surface versus structure on test case performance. The hypothesis being tested here is that analogy transfer, as measured by case performance (especially diagnostic and therapeutic accuracy), will be improved if the reasoner is led to focus on structural similarities between analogous cases. Figure 2 graphically depicts the "easy" versus "hard" case orderings.

In short, all 11 cases (including 6 study analogs, 3 nonanalogous distracter cases, and 2 test cases) were identical among all subject groups. However, depending upon the specific fixed order in which subjects received the cases, juxtaposition of analogs with similar surface *and* structure versus only similar structure varied. It was then possible to test whether subjects were able to better elicit and assimilate structural characteristics (diagnosis deterministic) when *not* distracted by surface similarities, as in the "hard" case ordering. If so, then they should be led to investigate the appropriate diagnostic category and, therefore, more often attain diagnostic specificity (i.e., make an accurate specific diagnosis within the diagnostic category).

Subject Recruitment

To recruit medical students, announcements were made at two major U.S. medical schools, one in the southeastern United States and one in the western United States. A Web site was programmed that allowed students to enroll in MARS. The Web site provided basic information about MARS, as well as a Web form into which students could enter their name, e-mail address, mailing address (for remuneration purposes), and a desired password to the site. Although no

Figure 2. Case-ordering method. C = Surface features, where the numeric subscript indicates the type of characteristics and the letter indicates a particular instance of those characteristics. S = Structural features (subscript number indicates type of structure, letter indicates instance). D = Diagnostic category (i.e., upper respiratory infection versus CHF). In the "easy" case order, subjects received two beginning cases followed by two cases that match both the structure and the surface characteristics of the initial cases. These cases are followed by two harder cases that require cross-mapping; that is, the previous structurally analogous cases had different surface characteristics. In the "hard" case order, subjects received the harder cross-mapped cases first followed by the easier cases. Thus, *Source (C1aS1D1) > Target (C1bS1D1)* is an "easy" sequence because both the surface characteristics and structure of the target case match those of the source case. Conversely, *Source (C1aS1D1) > Target (C2aS1D1)* is harder because the surface characteristics are different between the source and target, and the same structure indicates that D1 is the correct diagnostic category. In a final test phase (cases 9 and 11), both groups receive cases with unique surface characteristics. Dashed lines indicate distracter cases (cases 1, 6, and 10) with no analogs.


"EASY" case ordering
"HARD" case ordering
Study Phase
Test Phase
1
2
C1aS1a
D1
3
C2aS2a
D2
4
C1bS1b
C2cS1c
5
C2bS2b
C1cS2c
6
7
C2cS1c
C1bS1b
8
C1cS2c
C2bS2b
9
C3aS1d
10
11
C3bS2d

verification of training level was performed, the announcement targeted senior (4th year) medical students, and students were asked to confirm their training status as 4th-year students on the enrollment form. Upon submission of the registration form, a server-side algorithm sequentially assigned each student into one of the study groups. Ninety-six (96) students enrolled in this manner.

Satisfaction Survey

The Satisfaction Survey consisted of six questions, four of which were asked of all subjects and two of which were asked only of those subjects in the analogy transfer evaluation (ATE) groups. The survey was made available to subjects as a simple Web form immediately upon completion of the final test case (Case #11) in their assigned sequence. Additional space was provided for free-text comments. The survey questions, with their Likert scale responses, appear in Table 2.

Additional MARS Results

Satisfaction Survey Results

Seventy-one (71) subjects completed the Satisfaction Survey, including 33 ATE group subjects and 38 control group subjects. In response to Question 1, regarding experience with case-based learning (CBL) tools, both groups rated an average 2.18 out of 4, indicating that they rarely used such tools. In response to Question 2, which rated Medulator™ as a case-based assessment tool, ATE subjects rated 4.09 and control subjects rated 4.29 (cumulative rating 4.20) out of 5,

Table 2. MARS Satisfaction Survey questions (Questions 1 to 4 were asked of all subjects, and Questions 5 and 6 were asked only of ATE subjects.)

Question	Response Options
1. To what extent have you used other case-based tools similar to Medulator™?	1 = Never 2 = Rarely (q year or less) 3 = Occasionally (q month or less) 4 = Often (several times a month)
2. How would you rate Medulator™ as a case-based assessment tool?	1 = Poor 2 = Fair 3 = Average 4 = Very good 5 = Excellent
3. How applicable is Medulator™ as a potential supplement to your training?	1 = Not at all 2 = Limited 3 = Somewhat 4 = Highly 5 = Completely
4. How would you rate the difficulty of the study cases, in general?	1 = Too easy 2 = Fairly easy 3 = Neither hard nor easy 4 = Moderately difficult 5 = Too difficult
5. How effective has Medulator™ been at helping you to think "analogically" (i.e., to understand new cases better by mentally retrieving the solution to previous cases)?	1 = Totally ineffective. I still don't understand the concept or benefit of analogical reasoning. 2 = Little effect. I understand the concept but not the benefit of analogical reasoning. 3 = Somewhat effective. I understand the concept and believe it may be beneficial. 4 = Very effective. I clearly understand the concept and benefit.
6. To what extent did using analogical reasoning help you to solve cases?	1 = No help. In fact, it hindered me. 2 = It neither helped nor hindered me. 3 = Somewhat helpful. I could apply the concepts from previous cases to solve new cases. 4 = Very helpful. I was a more effective diagnostician because of analogical reasoning.

indicating that both groups, on average, thought highly of the program. Overall, 93% of respondents rated 4 (very good: 46 respondents) or 5 (excellent: 20 respondents). In comparison, ATE subjects and control subjects rated applicability of Medulator™ to their training (Question 3) at 3.76 and 3.92, respectively (cumulative rating 3.85 out of 5). Overall, 73.2% of respondents rated 4 (highly applicable: 40 respondents) or 5 (completely applicable: 12 respondents). User comments (see Table 3) suggested that applicability may have suffered due to the heavy use of bioterrorism cases. Regarding case difficulty (Question 4), ATE subjects rated 3.82 and control subjects rated 3.92 (cumulative rating 3.87) out of 5. Overall, 85.9% of respondents rated case difficulty at 4 (moderately difficult: 59 respondents) or 5 (very difficult: 2 respondents), suggesting that the cases were perceived as moderately difficult in general.

For ATE subjects only, an average rating of 3.12 out of 4 was given for Question 5, indicating that Medulator™ was perceived as being at least somewhat effective at helping these students think "analogically." Overall, 97% of respondents rated 3 (somewhat effective: 27 respondents) or 4 (very effective: 5 respondents). Also, ATE subjects gave an average rating of 2.88 out of 4 for Question 6, suggesting that using analogical reasoning was perceived as being somewhat helpful to subjects' ability to solve cases. Overall, 75.8% of respondents rated 3 (somewhat helpful: 21 respondents) or 4 (very helpful: 4 respondents).

A sample of user comments appears in Table 3. In general, user comments fell into three categories: (1) comments regarding Medulator™ as a learning tool,

Table 3. Specific user comments

"This is a very effective and useful way to learn medicine. I wish that there were more programs like this at use in my school."
"I really liked the interface. The program is easy to use and self-explanatory. I learned a lot and feel that this would be very useful in medical school."
"I thought this was a really great learning tool, not just because of the analogous nature, which I loved, but also the way the program was set up—easy to use, intuitive, etc."
"The cases were really good. I think that more common diseases are important (not q fever or anthrax) for medical education. I enjoyed the cases more than the analogical reasoning just because it seemed to get tedious and I had trouble remembering each case."
"I quite liked the format of the Medulator™; however, I thought the content was pretty obscure and full of zebras. What's up with all the explosions and crop dusting?" [Note: This subject rated 3 out of 5 for applicability.]
"This is such a great Medulator™! Best I have ever seen! Although it seems that the study cases lean more heavily towards hazardous exposures (cyanide poisoning, bioterrorism act, etc.)."
"I think this would be an excellent program if the cases were more common medical problems. There were some cases like that, but I felt I had a lot of weird pneumonias and biochemical agents. I am not going to see that in my everyday practice." [Note: This subject rated 3 out of 5 for applicability.]
"This is a very effective tool to get much of the benefit of seeing multiple patients in a fraction of the time. I did feel that the analogical reasoning sections were not very helpful though. There are too many variables to make definite [sp] determinations about the similarities of various cases."
"The idea of analogical reasoning was useful in solving cases by seeing the analogies with previous cases. . . . I believe that the format is very useful as a study/learning tool but could be more useful if it focused on diseases/problems that are more commonly seen in hospitals."

(2) comments regarding the bioterrorism nature of some cases, and (3) comments regarding the analogical reasoning component. It should be noted that Medulator™ was operating closer to pure assessment mode during MARS because context-specific feedback was limited, except during the ATE exercise and for the Final Diagnoses and Final Treatments assessment sections. It is also apparent that subjects did not like the bioterrorism focus of the four respiratory infection cases. Furthermore, respondents who commented negatively on the bioterrorism cases also rated the applicability of Medulator™ to their training lower than the group average. Finally, some respondents offered specific reasons why they believed that the analogical reasoning component was not as helpful, including difficulty in remembering cases and "too many variables."

CHAPTER 4

FINAL DISCUSSION AND CONCLUSIONS

This thesis described the development, modification, and implementation of Medulator™, a commercially available online patient simulation program created for the teaching and assessment of medical students and practicing medical professionals. The program was used to study the effects of several important principles of learning, as suggested by cognitive psychology research, on the clinical reasoning of novice clinicians (i.e., medical students).

To test the hypothesis that encouraging reasoners to focus on the structural features of a target case would lead to improved knowledge transfer, we manipulated the surface characteristics and structure of cases using different combinations and different instances of each via varied case orderings. However, the only finding of statistical significance in this experiment was an interaction between case order and analogy transfer evaluation (ATE) with respect to correct treatment selection. As hypothesized, when subjects were taught to focus on case structure in the harder case ordering, where surface and structure were *not* aligned in early analogs, performing explicit case comparisons led to improved therapeutic accuracy. However, diagnostic accuracy was not affected under the same conditions. Furthermore, the positive effect on therapeutic accuracy was primarily driven by cases in the more familiar knowledge domain (cardiology), which appears to indicate that case-based reasoning works better for diagnoses that medical students have previously learned. This would seem to be consistent with the domain-specificity concept introduced by Elstein [1].

It is also possible that in the more familiar cardiology cases the ATE exercise may be resulting in analogical reminding from a schema of this type of case (CHF), which the subjects may have previously encountered (before Medulator™ training), a process that may be assisted when they are focusing on structural comparisons due to the "hard" case ordering. In the less familiar bioterrorism cases, there is no opportunity for analogical reminding because the subjects have likely never encountered similar cases before Medulator™. This is also a valid argument for using didactic methods that more passively transfer core clinical concepts before moving on to patient simulations, particularly when introducing new diseases. Intuitively, patient simulations in the purest sense provide the ability to practice applying previously learned concepts and should *supplement* rather than *supplant* lectures and presentations as the primary knowledge transfer media. A possible exception is patient simulation that operates in teaching mode and provides a significant, if not comprehensive, coverage of clinical concepts needed to address the "patient's" disease. Medulator™ can operate in such a teaching mode, but for obvious reasons related to the research objectives, full teaching mode was not used in MARS.

Also based on this study's findings, requiring explicit case comparison via the ATE exercise independent of case ordering appears to offer no benefit for case performance, with the exception of a trend toward lower work-up costs in the ATE group. In other words, diagnostic and therapeutic accuracy improved with successive cases with no significant difference between study groups. It would seem then that learning occurred through practice. However, additional experimentation using counterbalancing to control for possible differences in individual case difficulty would be required before such a conclusion could be deemed legitimate. Additional research would also be required to confirm the possible effect of ATE on test costs that was observed in MARS.

On the other hand, and in concordance with prior research findings [2-5], MARS has demonstrated clear benefits from the use of case summary. The incorporation of a case summary exercise within local case-based learning (CBL) methodologies using patient simulations would emulate the traditional practice of writing a History & Physical (H&P) summary or patient progress note ("SOAP" note). Admittedly, H&Ps and SOAP notes serve important communication and medical-legal functions, as well as a reminder for the clinician who writes them. Nevertheless, it is the reminder function of case summary that provides value to CBL application users, and it seems desirable to extract the most value from a particular CBL application by providing such a simple, yet effective, tool that the application user can employ as a self-reminder and proxy for self-explanation.

Finally, it is of interest that the ATE subjects perceived, on average, that using analogies improved their ability to solve cases, although indicators of knowledge transfer did not differ significantly between study groups. It may be that

ATE subjects felt as if they were being assisted by the ATE exercise yet were insufficiently influenced, intellectually unable, or otherwise disinclined to carry analogical principles forward to target cases. Perhaps by learned habit, subjects ultimately relied instead upon the more familiar pathophysiologic basis of disease approach to diagnostic reasoning. After all, it is unreasonable to think that brief manipulation of these student subjects, within the greater context of their medical training, could have more than a minor, transient impact on their reasoning habits. However, until a significant impact is demonstrated, it is unlikely that educators will be compelled to adopt analogical reasoning exercises as a prominent curricular component.

In conclusion, the purpose of this research was to identify new CBL methodologies that could enhance existing and future patient simulation applications. MARS may hold potential implications regarding the clinical reasoning processes of novice clinicians (e.g., medical students), with possible differences between diagnostic reasoning and therapeutic reasoning, and with differences between familiar and unfamiliar knowledge domains. Beyond this, the practical value of the ATE exercise is questionable, given the mixed results obtained in MARS. For example, surface and structural similarities between cases in real-world clinical practice occur at random or not at all, but not in "hard" or "easy" order. Nevertheless, if a strong association between explicit case comparison and diagnostic accuracy had been achieved, for example, then an argument could be made for the adoption of an ATE-like exercise. It is also intuitive and may still be reasonable to posit that the introduction of exercises that emphasize the causal and functional relationships among cases of similar diagnostic domain could have value in the medical curriculum. CBL is becoming a prominent pedagogical method in medical education. Therefore, the use of explicit-case comparison exercises should be considered in the CBL approach.

Additional research that also seeks to eliminate many of the limitations of MARS, as outlined in the published manuscript [6], is warranted. This is especially true since MARS failed to reproduce the clear benefits of case comparison demonstrated by others such as Gentner et al. [7] and Yanowitz [8-9] in nonmedical domains. For example, a study using experienced clinicians rather than novices may be more successful at demonstrating a positive effect from an ATE-like exercise, for reasons already mentioned. However, if it is true that experienced clinicians use analogical reasoning extensively, then I would speculate that such a study could again fail to demonstrate a difference between groups, precisely because the use of analogies is superfluous and the ATE effect would be transparent. Thus, a more fruitful study might compare the performance of novices and experts with respect to diagnostic versus therapeutic outcomes when using explicit analogical techniques.

Furthermore, it is worth a word of caution to future researchers by reiterating that it may be the simplicity of the previous studies' methods [7–9], or the complexity of medical cases themselves, or the interactive nature of patient simulations that preclude the ability to reproduce those results. There is necessarily a great leap in the cognitive process from the study and comparison of brief narrative scenarios to the interactive information gathering and assimilation required of realistic patient simulations (like Medulator™ cases), much less real-world clinical problem solving. Although it may be possible to simplify the medical scenarios used in order to demonstrate benefits from explicit-case comparison, the generalization of any results achieved to other patient simulation programs, much less real-world clinical practice, would be unrealistic and, therefore, less meaningful.

As one important outcome of this research, all Medulator™ cases now include case summary as a standard feature, and the use of an ATE exercise for analogous case sets is an optional feature. There is little dispute that patient simulation cases such as Medulator™ are valuable and popular components of CBL programs. However, to optimize the pedagogical impact of CBL, a continued search for worthy case-based support features is warranted. It also remains for future research to show whether CBL features that attempt to induce analogical reasoning in medical novices are worthwhile. Indeed, the question of how much the acquisition of any expert reasoning skills can be accelerated remains to be answered. Until then, effective analogical reasoning may remain a hard-earned characteristic of the expert clinician.

References

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

[2] Chi, M.T.H., Deleeuw, N., Chiu, M.H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439-477.

[3] Bielaczyc, K., Pirolli, P.L., & Brown, A.L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction, 13*(2), 221-252.

[4] Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 5,* 145-182.

[5] Neuman, Y., & Schwarz, B. (1998). Is self-explanation while solving problems helpful? The case of analogical problem solving. *British J Educ Psychol, 68,* 15-24.

[6] Thompson, G.A., Morrison, R.G., Holyoak, K.J., & Clark, T.K. (2006). Evaluation of an online analogical patient simulation program. *Proceedings of the Nineteenth IEEE International Symposium on Computer-Based Medical Systems* (pp. 623-628). Los Alamitos, CA: IEEE Computer Society.

[7] Gentner, D., Lowenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *J Educ Psychol, 95,* 393-408.

[8] Yanowitz, K.L. (2001). Using analogies to improve elementary school students' inferential reasoning about scientific concepts. *School Science and Mathematics, 101*(3), 133-142.

[9] Yanowitz, K.L. (2001). The effects of similarity of theme and instantiation in analogical reasoning. *Amer J Psychol, 114*(4), 547-567.