

The Medulator Analogical Reasoning Study (MARS)

Conducted in cooperation between

The University of California, Los Angeles, Dept. of Psychology

And

Medantic Technology, LLC

This document prepared by

Gregory A. Thompson, MD

Copyright © 2005, MD Informatics, LLC

BACKGROUND

Case-based reasoning is ubiquitous in real-world medical diagnosis, and 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** 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. Analogical reasoning involves comparison of structured information between two cases and allows the reasoner to make inferences about one case (the *target*) based on prior knowledge of another case (the *source*). Numerous laboratory studies have suggested conditions that may facilitate transfer via analogy (see 11-13 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 definition analogy involves comparison of cases based on structure (i.e., a comparison of the pattern of relations present in each case). For example, the diagnoses for two patients may be said to be analogous if they have similar patterns of symptoms and 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 race, ages or have a similar occupation). These non-diagnostic surface characteristics can frequently be quite salient and can distract reasoners from a full appreciation of the structural similarities between two cases. Thus, 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 (14); however, the presence of these surface similarities may in some circumstances lead the learner to overlook the diagnostic structural characteristics (15). 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 (16-18). For example, Gentner, Lowenstein and Thompson (18) 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 also 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 not be as sensitive to explicit comparison enhancement as 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 have modified the **Medulator** Final Assessment section to include an explicit Analogy Transfer Evaluation (ATE). The ATE requires case users 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 (4-7). In the context of ATE, Case Summary serves as a self-explanation proxy. Thus, a separate arm of this study would examine the effect of student Case Summary on case performance, independent of ATE.

The research described was designed and conducted in cooperation with Dr. Keith Holyoak from the UCLA Department of Psychology and Dr. Robert Morrison from Xunesis.

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 through 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 diagnostic process

METHODS

We used **Medulator** to study the effect of case ordering, explicit case comparison, and writing case summaries on diagnostic and treatment performance using cases that systematically varied structure (i.e.,

diagnosis determinants such as full symptom constellation, physical examination findings, diagnostic test results, response to therapy, etc.) and surface characteristics (i.e., salient, non-diagnostic information such as patient age, gender, occupation, chief complaint, presenting symptoms, etc.). Participants were senior medical students.

Each participant 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 (Hypertensive CHF, Idiopathic Dilated Cardiomyopathy with CHF, Acute MI with CHF, and Infective Endocarditis with CHF) and (3) three non-analogous distracter cases. Study cases were structurally analogous within their own diagnostic category but superficially similar within and/or across diagnostic categories. Cases were presented in two different sequences: (1) “easy” analogies (surface and structure matched) followed by “hard” analogies (only structure matched), and (2) hard analogies followed by easier analogies.

Ninety six (96) senior medical students who had never used **Medulator** self-enrolled and were paid \$150 for their participation which took 5.0 hours on average. Participants were randomly assigned to one of six groups (see Table 1). Participants completed cases in a defined order (see Figure 1). For Groups 1 and 4, explicit case comparison was evoked via Analogy Transfer Evaluation (ATE) on 2nd, 3rd, and 4th analogs of each analogous case set; in these two groups, students were encouraged to write a Case Summary to serve as a reminder of cases’ salient features. All other groups received no instructions for comparing previous cases but proceeded directly to the default Final Assessment (diagnosis and treatment selections, and post-test). However, Groups 2 and 5 wrote Case Summaries (without ATE) while Groups 3 and 6 did not.

The ATE condition consisted of 2 parts:

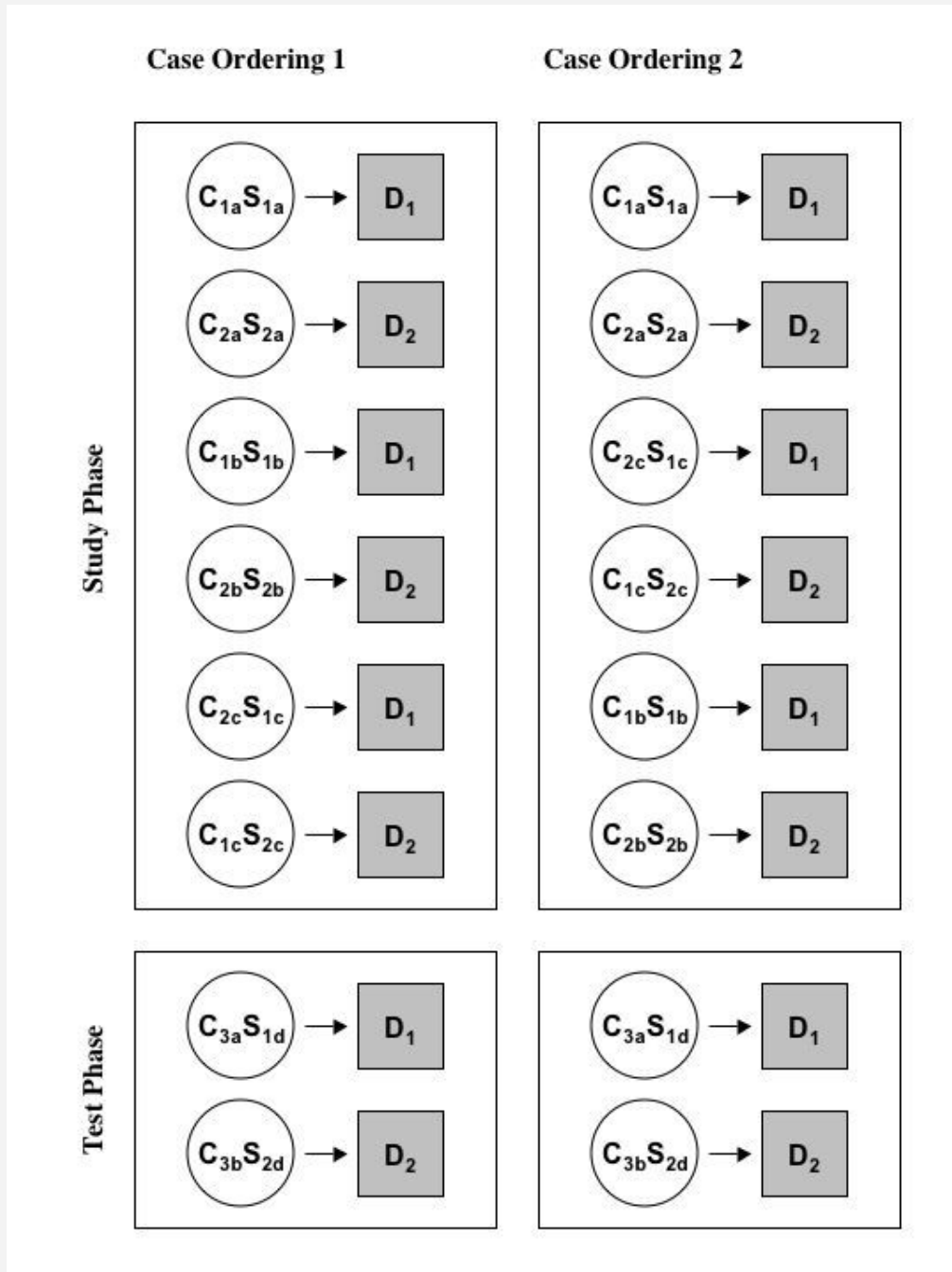
1. First, once students completed selecting their Final Diagnoses and Final Treatments, they were presented with instructions to rate the degree to which previously completed cases were structurally similar to the current case. Upon submitting their ratings, they were given feedback as to which case(s) 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, etc. 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.

Table 1. Design of Experiment

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

Figure 1. Design of Experiment. Both Groups received a series of cases for which they had to make a diagnosis. Participants received feedback after each case. Case Ordering Group 1 received two beginning cases followed by two easy cases that matched both the structure and the surface characteristics of the initial cases. These cases were followed by two harder cases that required cross-mapping (i.e., the previous structurally analogous cases had different surface characteristics.) Case

Ordering Group 2 received the last four with the harder cross-mapped cases first followed by the easier cases. In a final test phase, both groups received cases with unique surface characteristics. In the figure below C represents surface characteristics and S represents structure. Structure always matches the diagnosis type (D). The number after C or S represents a class of surface or structural characteristics and the letter following the number indicated a particular instance of that class.



Dependent Variables measured:

1. Number of treatment attempts required to achieve a positive patient outcome
2. Number of Correct Final Diagnoses chosen
3. Number of Incorrect Final Diagnoses chosen
4. Number of Correct Final Treatments chosen

5. Number of Incorrect Final Treatments chosen
6. Number of diagnostic tests chosen
7. Costs of diagnostic tests chosen
8. Final Quiz Score[†]
9. Case time (keyboard time)

[†] A 5 MCQ post-test was given after each case and was designed to test how well students recalled the surface characteristics of each case, rather than further testing them on their knowledge of disease.

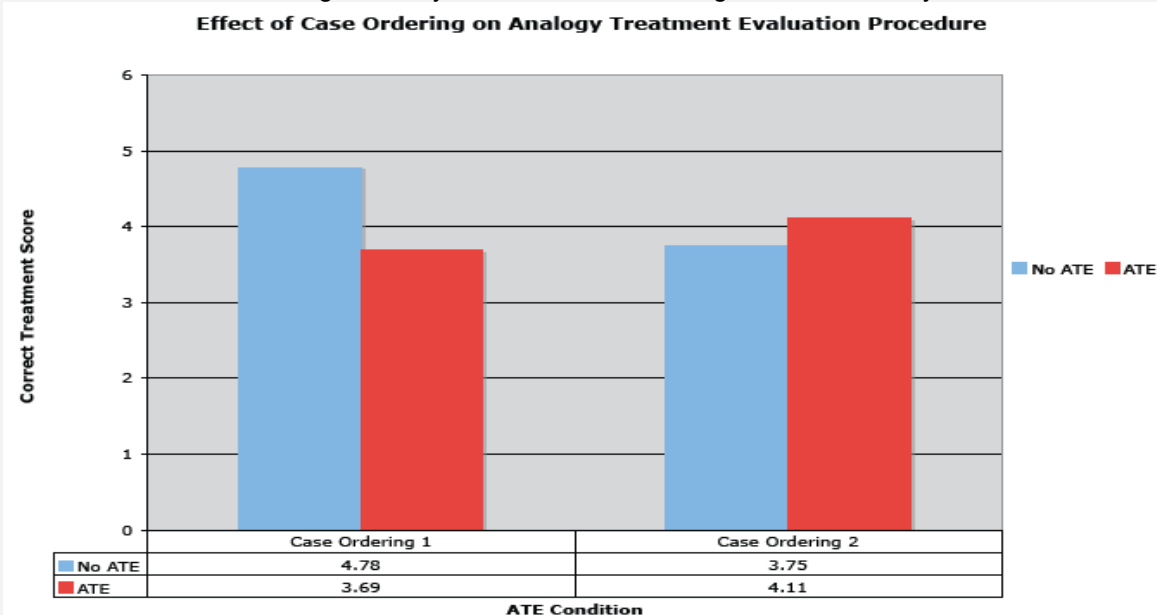
Participants were initially allowed 2 weeks to complete the study. However, in order to achieve the goal of at least 64 completions, some subjects were granted up to 3 one-week extensions. A MARS Honor System was published stating that subjects should work independently and with no external assistance. Each subject was asked to complete a wrap-up survey after completing all 11 cases.

RESULTS

Of the 96 subjects who originally self-enrolled, 72 subjects completed all 11 cases (33 in ATE group, 39 controls) and 71 students completed the wrap-up survey. 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 sequencing on performance we analyzed test cases 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 participants in this analysis wrote case summaries for every case. Half of the participants in this analysis compared cases using the ATE instructions (groups 1 and 4) while half just performed case summaries (groups 2 and 5). We performed a 2 x 2 between-subjects ANOVA to investigate the effect of case sequencing and explicit comparison (ATE) as well as their possible interaction. Though individual effects were not statistically significant for either case sequencing alone, a reliable cross-over interaction was seen ($p=.009$), as shown in Figure 2. On closer inspection of the data the interaction appears to be driven by the cardiology case and not the bioterrorism case. There were no other reliable effects for other dependent variables in this analysis.

Figure 2. KEY: Case Ordering 1 = Easy to Hard; Case Ordering 2 = Hard to Easy



Explicit Comparison (ATE). To evaluate the effectiveness of explicit comparison between cases as implemented using the Analogy Transfer Evaluation (ATE), we only included cases in which the participants in the ATE group were given ATE instructions (thus the first cases of each diagnosis as well as the unrelated cases were eliminated). Groups who did not do case summaries (Groups 3 and 6) were not included in this analysis. We performed a mixed 2 x 2 ANOVA which factored diagnostic category (i.e., Cardiac vs. Bioterrorism cases) as a within-subject factor and explicit comparison (ATE vs. no-ATE)

as a between subject variable. Participants spent more on tests (i.e., test costs) on Bioterrorism cases ($p=.03$) and there was a trend ($p=.05$) toward a reduction in test costs in the ATE condition. Likewise, Diagnosis, Treatment, and Quiz performance were all 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, independent of ATE, we looked at all Cardiac 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 which factored diagnostic category (i.e., Cardiac vs. Bioterrorism cases) as a within-subject factor and whether or not participants were required to write case summaries as a between subject variable. Diagnostic accuracy reliably increased ($p<=.002$) while treatment attempts decreased ($p=.05$) when participants were prompted to write case summaries. A trend suggested that writing case summaries improved diagnosis of Bioterrorism cases more than Cardiac cases ($p=.06$) and treatment attempts mainly decreased with case summary writing for Cardiac cases ($p=.04$). Writing case summaries had no reliable effect on test costs, treatment score, or quiz score.

DISCUSSION

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

Though results from this study were mixed, there was not strong evidence that applying explicit comparison of cases using ATE had a positive influence on most measures of performance. Likewise case ordering did not have a reliable effect on test case performance. There is a suggestion in our results, however, that explicit comparison can improve performance when students encounter difficult cases first, particularly when they are somewhat familiar with the basic principles involved in the diagnostic category. 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 non-diagnostic surface characteristics. In the “easy” case ordering surface and structural characteristics were aligned across early cases and thus students may have mistakenly associated surface characteristics with diagnostic efficacy. In contrast, when non-diagnostic surface characteristics do not align with diagnostic structural characteristics in early cases, explicit comparison via the ATE seems to improve later test case performance. This appears to be particularly true for cases in which the students may already be somewhat familiar with the treatment principles (Cardiac vs. Bioterrorism cases). Thus, this suggests that for new areas of medical learning it is necessary to first educate students on important diagnostic and treatment principles before moving on to case-based education methods.

There are several possible reasons why explicit comparison (ATE) did not improve subjects' case performance. First, it is possible that the students participating in this study did not have the necessary expertise to abstract the diagnostic and treatment principles from the cases. Thus, explicit comparison of cases did not reinforce these principles (see Chi, Feltovich, & Glaser (10) for a similar issue in the domain of physics). This hypothesis is consistent with the results in treatment score performance noted in the previous paragraph.

A second, alternative explanation is that medicine, like law, and unlike business is an intrinsically analogical domain and thus fourth year medical students (in contrast to Gentner, Lowenstein and Thompson's (18) business students) instinctively think analogically. Thus forcing explicit comparison of cases is superfluous.

A third reason 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 yet been attempted. While this explanation is possible, medical professionals obviously use very old knowledge from previous cases to diagnose new cases. Thus, this explanation may interact with the experience of the student/professional.

Lastly, 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 superficial versus structural characteristics (e.g. 18). Also, in those studies, information was passively transferred, such that subjects were insured of being exposed to all structural characteristics germane to solving the problem. In contrast, cases which 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 between cases. Subjects could be overwhelmed when trying to determine which structural categories of data (historical, diagnostic, therapeutic, etc.) 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 a number of other limitations of the present study. Cases were not standardized for level of difficulty prior to the study. However, all cases were authored by a Board-certified Internal Medicine physician, reviewed internally by a separate Board-certified physician, and were deemed to be of equal (or similar) difficulty level within diagnostic categories. That is, all Bioterrorism cases were written to be similarly challenging, as were all Cardiac cases. Bioterrorism cases were expected to be (and indeed proved to be) more difficult than Cardiac cases.

Human factors could have led to overestimation of some dependent variables in both groups. For example, subjects cannot advance to the Final Assessment section of a case until they achieve a positive patient outcome (based upon selected treatment regimens). However, because subjects were not required to advance to the Final Assessment section of each case upon achieving a positive patient outcome, they could have continued to select additional treatment regimens in order to observe their effect on patient outcome. While such “curious experimentation” may be valuable from a learning standpoint, it leads to overestimation of treatment attempts. Likewise, 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 assure a correct selection, leading to spuriously high numbers of incorrect selections.

Also, as with any testing where students are allowed to work independently, we were unable to control potential seeking of assistance from informational resources (e.g. Web, textbooks, other students) beyond those provided internally within the cases themselves.

An additional 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 one week or less (19). 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 meticulous in their 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 drop-out), study account extensions were granted in one-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 a month to complete all 11 cases, resulting in weeks between their first and last cases.

Of note, and of particular interest to designers of CBL programs, is case performance improvement resulting from the use of Case Summary 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 participants 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. While this result is consistent with previous research, 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 CBL platforms which incorporate such a simple yet effective tool. Furthermore, informal analysis of subjects’ case summaries written for this study revealed that even subjects in groups which did not perform ATE (and, therefore, did not need case summaries as a comparison tool) wrote word counts comparable to ATE subjects, suggesting that those subjects were using the Case Summary utility for reflection and self-explanation.

Finally, the follow-on survey results demonstrate a high degree of user acceptance of *Medulator* (see Appendix A for full survey results and interpretation). 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. Interestingly, the applicability of bioterrorism cases to subjects’ training was rated low.

REFERENCES

1. Fordis, M. et al. Comparison of the Instructional Efficacy of Internet-Based CME With Live Interactive CME Workshops, A Randomized Controlled Trial. *JAMA*, September 7, 2005—Vol 294, No. 9
2. Johnson, M. 2002. Introductory biology online. *J College Science Teaching* 31, 312-317.
3. Wisher RA, Olson TM. Research Report 1802 – The Effectiveness of Web-based Training. US Army Research Institute for the Behavioral and Social Sciences. January, 2003.
4. 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.
5. 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.
6. Neuman, Y., & Schwarz, B. (1998). Is self-explanation while solving problems helpful? The case of analogical problem-solving. *British Journal of Educational Psychology*, 68, 15-24.
7. Chi, M. T. H., Deleeuw, N., Chiu, M. H., & Lavancher, C. (1994). Eliciting Self-Explanations Improves Understanding. *Cognitive Science*, 18(3), 439-477.
8. Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, 12, 306-355.
9. Rattermann, M. J., & Gentner, D. (1987). Analogy and similarity: Determinants of accessibility and inferential soundness. In *Proceedings of the Ninth Annual Conference of the Cognitive Science Society* (pp. 23-35). Lawrence Erlbaum Associates, Mahwah, JN.
10. R 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.
11. Holyoak, K.J., & Thagard, P. (1995). *Mental leaps: Analogy in Creative Thought*. MIT Press: Cambridge, MA.
12. Gentner, D., Holyoak, K.J., & Kokinov, B.N. (eds) (2001). *The analogical mind: Perspectives from cognitive science*. MIT Press: Cambridge, MA.
13. 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.

14. Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development*, 67, 2797-2822.
15. Goldstone, R.L., & Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cognitive Psychology*, 46, 414-466.
16. Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15, 1-38.
17. Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 1147-1156.
18. Gentner, D., Lowenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*. 95, 393-408.
19. Wharton, C.M., Holyoak, K.J., & Lange, T.E. (1996). Remote analogical reminding *Memory & Cognition* 24: 629-643 .
20. Shim B, Brock D, Jenkins L. (2005). Developing practical criteria for evaluating online patient simulations: a preliminary study. *Med Teach*, 27(2):175-7.
21. Thomas MD, O'Connor FW, Albert ML, Boutain D, Brandt PA. (2001). Case-based teaching and learning experiences. *Issues Ment Health Nurs*. 22(5):517-31.
22. Schwartz RW, Donnelly MB, Nash PP, Johnson SB, Young B, Griffen WO Jr. (1992). Problem-based learning: an effective educational method for a surgery clerkship. *J Surg Res*. 53(4):326-30.
23. Neistadt ME, Wight J, Mulligan SE. (1998). Clinical reasoning case studies as teaching tools. *Am J Occup Ther*. 52(2):125-32.
24. 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):1050-5.
25. Shaw MJ, Beebe TJ, Tomshine PA, Adlis SA, Cass OW. A randomized, controlled trial of interactive, multimedia software for patient colonoscopy education. *J Clin Gastroenterol*. 2001 Feb; 32(2):142-7.
26. Maag M. The effectiveness of an interactive multimedia learning tool on nursing students' math knowledge and self-efficacy. *Comput Inform Nurs*. 2004 Jan-Feb;22(1):26-33.
27. Ross GC, Tuovinen JE. (2001). Deep versus surface learning with multimedia in nursing education development and evaluation of WoundCare. *Comput Nurs*. 19(5):213-23.
28. Wentling TL, et al. (2000). E-Learning – A Review of Literature. Available: <http://learning.ncsa.uiuc.edu/papers/elearnlit.pdf>
29. Hmelo, CE. (1990). Computer-Assisted Instruction in Health Professions Education: A Review of the Published Literature. *J Ed Tech Sys*. V18, n2, p83-101.
30. Sakowski HA, Rich EC, & Turner PD. Web-based case simulations for a primary care clerkship. *Academic Medicine* 2001; 76:547.
31. Leong SL, Baldwin CD, Adelman AM. Integrating Web-based computer cases into a required clerkship: development and evaluation. *Acad Med*. 2003 Mar;78(3):295-301.
32. Eshach H, Bitterman H. From case-based reasoning to problem-based learning. *Acad Med*. 2003 May;78(5):491-6.
33. Swagerty D Jr, Studenski S, Laird R, Rigler S. A case-oriented web-based curriculum in geriatrics for third-year medical students. *J Am Geriatr Soc*. 2000 Nov;48(11):1507-12.
34. Jonassen DH. Scaffolding Diagnostic Reasoning in Case-Based-Learning Environments. *J Comp High Ed*. 1996; 8(1), 48-68.