The Medulator Comparison of Methods Study (MedCoMS)

Conducted in cooperation between

The University of Utah, Department of General Internal Medicine

And

Medantic Technology

This document prepared by

Gregory A. Thompson, MD

Medulator Comparison of Methods Study (MedCoMS)

BACKGROUND

E-Learning has been shown to compare favorably to traditional teaching methods as measured by standard achievement tests and learner satisfaction.(1,2,3,28) And though there is a paucity of such studies demonstrating that the increased knowledge gained is of greater practical use, one recent example is provided by Fordis et al. (2005) where online CME was shown to effect changes in physician practice behavior better than traditional (live) CME.(1) As well, an online review of medical and educational literature databases (Medline, ERIC, and Cinahl) revealed a paucity of investigations which objectively compare pedagogical methods in medical education using an online CBL component to traditional methods. Practical criteria for evaluating online patient simulations are only recently being proposed.(20) Sakowski, Rich, and Turner (2001) found that students using Web-based case simulations as an individual exercise did not perform differently on the clerkship written examination than those in the traditional clerkship curriculum.(22) However, this study used only a small number of subjects and cases were not truly interactive, because no user input was elicited and no feedback was given. Leong, Baldwin, & Adelman (2003) compared different methods of case delivery and concluded that computer cases are better than paper cases or simple study articles.(23) And Swagerty et al. (2001) showed pre-test to post-test improvement in third year medical students enrolled in a case-oriented Web-based curriculum. Other examples of unilateral testing of case-based courses exist. But overall, few relevant studies can direct our investigation to detect differences based on pedagogical format. Making comparisons even more difficult are the widely various formats – from paper-based tutorials to medium fidelity interactive computer-based simulations to the most sophisticated haptics-based virtual reality programs – that qualify as "patient simulations."

Still, CBL is universally well-received by students, and the use of online CBL is burgeoning in medical schools. Literature espousing the effectiveness of Problem-Based Learning and it's core component CBL is readily available in a broad array of healthcare education settings.(21,22,23,24) And evidence of the benefits of interactive multimedia, including its ability to promote "deep" rather than "surface" learning, is likewise abundant, dating back to before the use of personal computers.(25,26,27)

Computer-based interactive patient simulations can help students affect transfer of 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 patient problems.(34) Thus, in theory, supplementing didactic information delivery in medical education curricula with an online CBL platform that is rich in interactive multimedia should improve knowledge retention through practice of knowledge application and deep memory formation. However, 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. It remains to be seen if learning can occur through highly interactive stand alone case simulations when used to reinforce learned concepts – this is the focus of MedCoMS.

A major advantage of the ***Medulator*** Virtual Patient case platform is that it is entirely data-driven, with the ability to track highly specific aspects of user performance (e.g. diagnostic and therapeutic accuracy, cost-effectiveness of patient evaluation). Thus, in collaboration with the University of Utah's Departments of Internal Medicine and Clinical Epidemiology, MT investigators designed and conducted a study comparing a ***Medulator***-based course on upper respiratory infections with traditional classroom delivery of similar content.

Research Objective

To compare pedagogical effectiveness of an online case-based course to that of traditional classroom-based delivery, using ***Medulator*** performance data in addition to pre-test and post-test performance.

METHODS

- 3 faculty members at the University of Utah School of Medicine:
  - Wrote teaching material addressing 2 different upper respiratory infections: Acute Rhinosinusitis and Acute Bronchitis
  - Provided and managed housestaff completion of study requirements
- Content consisted of:

- 12-question paper-based multiple-choice pre-test.
  - Classroom material: Two 20+ minute PowerPoint-based lectures, delivered sequentially.
  - Online material: Two 10-minute Flash-based didactic presentations, using the verbatim classroom-based lecture transcript, plus 2 ***Medulator*** Virtual Patient cases which were stripped of some teaching features in order to reduce time-on-task.
  - Test material: Three ***Medulator*** Virtual Patient cases, with all teaching points and case discussion removed (ie. pure summative mode), plus a 10-question online multiple-choice post-test, covering the material presented on Acute Rhinosinusitis and Acute Bronchitis.
- 139 Internal Medicine residents and interns were randomized to one of 2 study groups:
  - Control group: Receives classroom lectures
  - Intervention group: Completes online study material
- No subject had previously used ***Medulator***.
- Control (classroom) group subjects sat in-class on one of 3 lecture delivery dates, at the beginning of which they completed the 12-question paper-based MCQ pre-test.
- Intervention (online) subjects completed the same pre-test individually, and then were given instructions for accessing and completing the online material (didactic presentations + 2 study cases).
- Following a minimum 10-day waiting period after completion of either classroom lectures or online study material, all subjects were given instructions for accessing the online test case material.
- Performance data was computer-generated during online test case sessions, and consisted of the following dependent variables:
  - Number of treatment attempts per case to achieve a positive patient outcome
  - Number of diagnostic tests ordered and costs incurred, per case
  - Number of correct and incorrect Final Diagnoses per case
  - Number of correct and incorrect Final Treatments per case
  - Case (keyboard) Time per case
  - Quiz Score

Statistical Methods

Groups were compared on unordered categorical variables with two outcome categories using the chi-square or Fisher's exact test as appropriate. For unordered categorical variables with more than two categories, the chi-square test or Fisher-Freeman-Halton test was used as appropriate. For ordered categorical variables, the groups were compared using the Wilcoxin-Mann-Whitney test. Groups were compared on continuous variables using the two-sample *t* test, or the Wilcoxon-Mann-Whitney test if the distributions were skewed, as indicated by a significant Shapiro-Wilks test computed separately for each group. For skewed distributions, the median and interquartile range was used to summarize the distribution, rather than the mean and standard deviation. For multiple continuous outcomes, such as quiz scores, the average of the multiple outcomes was used as the outcome measure.

To adjust for pretest test scores using an analysis of covariance approach, linear regression was used if the outcome variable was sufficiently normally distributed, using the pretest score as a covariate. For skewed outcome variables, ordinal logistic regression was used if the outcome had a narrow range and linear regression on a log-transformed outcome variable if it had a wide range.

All reported *p* values are for a two-sided comparison, computed using Stata/SE 9.1 statistical software (StatCorp, College Station, TX).

RESULTS

Of the 139 subjects assigned to complete the materials, 129 subjects, representing training levels PGY1 thru PGY3, completed the study (72 online, 57 classroom). Dropout of 10 subjects occurred for a variety of reasons, including vacation, graduation from the residency program prior to completion, and otherwise failure to finish all test cases by study completion. Six subjects, originally randomized to the classroom group, failed to show for class and were re-assigned to the online group.

Online subjects spent an average of 15.25 minutes and 14.53 minutes on the two study cases (for a total average of 29.8 minutes of case time), and the two online didactic lecture presentations required a total of 20.2 minutes of runtime. Thus, average total time-on-task for the online group was 50 minutes. In comparison, total lecture time for the two faculty members delivering PowerPoint lectures in class averaged 45-50 minutes in each of the 3 classroom sessions. The average time from study session to testing was 16.4 days for the online group and 17.8 days for the classroom group.

To test the effectiveness of each instructional method, we measured each dependent variable provided by ***Medulator***, as well as performance on each of two 5-MCQ post-test quizzes. Results appear in Table 2.

Table 2. Main outcome results for 129 subjects who completed the required 3 test cases.

	Online [n = 72]	**Classroom [n = 57]**	**unadjusted *p* value**	**adjusted *p* value****
Post-Graduate (Residency) Year 1st 2nd 3rd 4th (recently graduated PGY3)	 30 (41.2) 15 (20.8) 25 (34.7) 2 (2.8)	 21 (36.8) 12 (21.1) 24 (42.1) 0 (0)	0.637	--
Pretest score, mean (SD) (high score is best)	61.7 (14.9)	56.9 (12.3)	0.056	--
Mean of two **Quiz Scores** (0 to 5 possible), median (25th , 75th percentile) (high score is best)	4.0 (3.5 , 4.5)	3.5 (3.5 , 4.0)	0.044	0.051
Mean **Case Time*** (time required to complete each case), minutes, median (25th , 75th percentile) (low score is best)	13.7 (10.3 , 17.3)	17.5 (13.3 , 23.2)	<0.001	0.001
Mean **Treatment Attempts** per case, median (25th , 75th percentile) (low score is best)	2 (1.4 , 2.7)	2.7 (2, 3.3)	0.002	0.005
Mean number **Correct Treatments** mean (SD) (high score is best)	3.8 (0.5)	3.4 (0.6)	<0.001	0.001
Mean number **Incorrect Treatments** median (25th , 75th percentile) (low score is best)	0.33 (0.00 , 0.67)	0.33 (0.00 , 0.33)	0.110	0.144
Mean number **Correct Diagnoses** mean (SD) (high score is best)	1.0 (0.2)	1.0 (0.2)	0.404	0.437
Mean number **Incorrect Diagnoses** median (25th , 75th percentile) (low score is best)	0.17 (0.00 , 0.33)	0.33 (0.00 , 0.33)	0.258	0.398
Mean number of **Tests Ordered**, median (25th , 75th percentile) (low score is best)	1.00 (0.67 , 1.67)	1.33 (0.67 , 1.67)	0.222	0.477
Mean **Test Costs** (cost of all tests ordered), $US median (25th , 75th percentile) (low score is best)	45 (20 , 71.7)	45 (38.3 , 62.8)	0.853	0.986

* Case times greater than one hour were set to missing before computing average of the three cases, the assumption being the student stepped away during case exercise.
** *p*-value adjusted for pretest score in a regression model.

The online group reliably took less time to complete each test case (p<.001), required fewer treatment attempts to achieve a positive patient outcome (p=.002), selected more correct treatment options (p<.001), and scored higher on the post-test quiz than the classroom group (p=.044). After adjusting for a non-significant difference in pre-test scores (p=.056), these effects remained reliable except post-test quiz scores showed a nearly reliable trend (p=.051).
There was no significant difference for other dependent variables, including incorrect treatments, correct and incorrect diagnoses, number of tests ordered, and test costs.

DISCUSSION

In this study we attempted a head-to-head comparison of instructional methods, with the main intervention being a dual-methodology approach of supplementing didactic material with interactive patient simulations. The dual-methodology platform was delivered via the Web versus a traditional approach of delivering the same instructional principles and concepts synchronously and podium-style to a classroom of subjects. Because online delivery of didactic material can be accomplished in a fraction of the time it takes human lecturers – the online didactic presentations used the verbatim lecture transcripts but took a total of 20.2 minutes, compared to 45-50 minutes for classroom delivery – we were able to augment the online didactics with ***Medulator*** Virtual Patient case studies and still match time-on-task to the traditional method. However, in order to accomplish this, many of the teaching features of ***Medulator*** were stripped for the study. Thus ***Medulator*** was operating much closer to pure assessment mode.

Our hypothesis that supplementing didactic material with interactive multimedia patient simulations would improve delayed performance proved largely true. The difference between the groups in time passed from study material to testing was only 1.4 days, which is not felt to be meaningful. However, despite the fact that the classroom group spent an average of 22% longer to complete their cases, the online group's performance was significantly better than the classroom group's in several of the most important performance indicators (treatment attempts, correct treatments, and quiz score).

At initial glance, it is not surprising that the online group took less case time, since they practiced on 2 study cases prior to the test phase. However, analysis of the differences in case time showed no reduction in the gap of case times from Test Case 1 (between-group difference of 2.2 minutes) to Test Case 3 (also 2.2 minutes), which would be expected if prior practice were the only explanation. Alternatively, it may be that the online group truly became more comfortable with the main principles and concepts to be learned in the study phase and were thus more efficient in applying those concepts in the test cases.

No significant difference was seen between the groups for dependent variables such as diagnostic accuracy or cost-effectiveness of diagnostic testing. The lack of effect with respect to diagnostic accuracy was to be expected given the narrow differential diagnosis that subjects were led to expect (acute viral sinusitis, acute bacterial sinusitis, or acute bronchitis). And the lack of effect with respect to cost-efficiency of diagnostic testing suggests that both groups understood the concepts presented equally well (e.g. appropriate indications for diagnostic testing such as chest x-ray for cough, and lack of efficacy of sinus radiography in cases of acute sinusitis).

There were a number of limitations in our study. 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 (and case time). 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.

Another limitation includes ***Medulator***'s method for measurement of case times, which were server-generated keyboard times. This method of time measurement tends to overestimate the time that users actually spend on case activity since they may step away for uncontrolled periods from an open browser, with the clock running.

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, and other students) beyond those provided internally within the cases themselves. However, the brief average case times – in other studies using more complex ***Medulator*** cases, case times averaged 25 - 30 minutes – suggest that any time taken to consult external resources was minimal or nil.

Finally, our population of subjects represents a single major Internal Medicine training program. Lack of information regarding this population's previous training and knowledge on the subject matter material, relative to other training programs, limits our ability to generalize our results.

Due to insufficient sample size, we were unable to measure an effect of the ***Medulator*** cases independent from the online didactic presentation component. Thus, the question remains if, and if so how much, ***Medulator*** improves performance over simply re-formatting live lecture material to the asynchronous Web environment. Previous studies have shown that simply re-formatting paper-based material to a more interactive Web version improves achievement scores.(31) However, our study approach and findings are consistent with the findings of Fordis et al. (2005), in which online CME, including a choice of multimedia presentations plus interactive cases with feedback, compared favorably to live CME activities and actually improved physician practice behavior.(1) However, the Fordis study also included both live and online forms of enabling tools and reinforcement that were not part of our instructional methods.

The cases used for this study were relatively simple inasmuch as the disease differential diagnoses, and therefore the therapeutic options, were limited. Because subjects knew that the cases covered recent lecture material on 2 specific disease states, albeit important Ambulatory Care disease states, they were conditioned for those diseases in the cases. It would be interesting in future research to use more complex cases, with broader differential diagnoses, and which are administered as part of a larger unstructured curriculum, in order to enable larger effect sizes in dependent variables such as diagnostic accuracy and diagnostic cost-effectiveness.

Nevertheless, we conclude that the Web-based approach adopted for this sample population, wherein online didactic material was supplemented with interactive cases, led to overall increased knowledge retention and ability to apply learned concepts in a computer-simulated teaching/assessment environment. ***Medulator*** is an effective way to reinforce such learned concepts in a benign, asynchronous, online practice environment, as well as to expose students and clinicians to more cases than they see during normal patient care, and to measure detailed aspects of their performance.

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. *Journal College Science Teaching* 31, p. 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: 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.