Studying the Effects of Personalized Language and Worked Examples in the Context of a Web-Based Intelligent Tutor

Bruce M. McLaren[1], Sung-Joo Lim[1], France Gagnon[2],
David Yaron[3], and Kenneth R. Koedinger[1]

[1] Human-Computer Interaction Institute, 5000 Forbes Avenue, Carnegie Mellon University, Pittsburgh, PA, 15213-3891 United States
bmclaren@cs.cmu.edu, sungjol@andrew.cmu.edu, koedinger@cs.cmu.edu
[2] Faculty of Education, University of British Columbia, Vancouver, BC V6T 1Z4
france.gagnon@shaw.ca
[3] Chemistry Department, 5000 Forbes Avenue, Carnegie Mellon University, Pittsburgh, PA, 15213-3891 United States
yaron@cmu.edu

Abstract. Previous studies have demonstrated the learning benefit of personalized language and worked examples. However, previous investigators have primarily been interested in how these interventions support students as they problem solve with *no other* cognitive support. We hypothesized that personalized language added to a web-based intelligent tutor and worked examples provided as complements to the tutor would improve student (e-)learning. However, in a 2 x 2 factorial study, we found that personalization and worked examples had no significant effects on learning. On the other hand, there was a significant difference between the pretest and posttest across all conditions, suggesting that the online intelligent tutor present in all conditions *did* make a difference in learning. We conjecture why personalization and, especially, the worked examples did not have the hypothesized effect in this preliminary experiment, and discuss a new study we have begun to further investigate these effects.

1   Introduction

In a recent book by Clark and Mayer [1], a number of principles were proposed as guidelines for building e-Learning systems. All are supported by multiple educational psychology and cognitive science studies. We were especially interested in and decided to experiment with two of the Clark and Mayer principles:

Personalization Principle One: Use Conversational Rather than Formal Style

Worked Example Principle One: Replace Some Practice Problems with Worked Examples

In contrast to most previous studies, however, we wished to test these principles in the context of a web-based intelligent tutoring system (ITS), rather than in a standard e-Learning or ITS environment or, as in even earlier studies, in conjunction with problems solved by hand. The key difference is that an intelligent tutoring system provides more than just problem solving practice; it also supplies students with con-

1

text-specific hints and feedback on their progress. In particular, we were interested in whether personalized language within an ITS and worked examples provided as complements to ITS-supported problems might improve learning beyond the gains from the ITS on its own with formal, impersonal feedback. Furthermore, because we are interested in delivering intelligent tutoring systems over the Internet, we were interested in whether we could show results with a web-based ITS deployed in a distance learning environment.

Until relatively recently, providing intelligent tutoring in an e-Learning environment was quite difficult to achieve, given the implementation complexity and computational overhead of the typical ITS. However, advances in ITS authoring tools, an important area of ITS research [2], have started to overcome this obstacle. For instance, authoring software developed in our laboratory, the Cognitive Tutor Authoring Tools (CTAT) [3], make it possible to deliver tutors on the web. CTAT builds on successful research and development of cognitive tutors, intelligent tutoring systems that have been shown to lead to significant improvements in learning of high school math [4] and are currently in use in over 2,000 schools across the U.S. Perhaps the most important contribution of CTAT thus far is its support for developing *example-tracing tutors*, tutors that can be rapidly built by demonstration but are constrained to single-problem use[1]. On the other hand, these example-tracing tutors exhibit behavior that is very similar to full cognitive tutors and are lightweight enough to be deployed on the web. Inspired by our newfound ability to rapidly deploy tutors to the web, our general aim is to explore how we can leverage CTAT and e-Learning principles to improve web-based learning.

The Clark and Mayer personalization principle proposes that informal speech or text (i.e., conversational style) is more supportive of learning than formal speech or text in an e-Learning environment. In other words, instructions, hints, and feedback should employ first or second-person language (e.g., "You might want to try this") and should be presented informally (e.g., "Hello there, welcome to the Stoichiometry Tutor! …") rather than in a more formal tone (e.g., "Problems such as these are solved in the following manner").

Although the personalization principle runs counter to the intuition that information should be "efficiently delivered" and provided in a business-like manner to a learner, it is consistent with cognitive theories of learning. For instance, educational research has demonstrated that people put forth a greater effort to understand information when they feel they are in a dialogue [5]. While consumers of e-Learning content certainly know they are interacting with a computer, and not a human, personalized language helps to create a "dialogue" effect with the computer. E-Learning research in support of the personalization principle is somewhat limited but at least one project has shown positive effects [6]. Students who learned from personalized text in a botany e-Learning system performed better on subsequent transfer tasks than students who learned from more formal text in five out of five studies. Note that this project did not

---

[1] In [3] these specialized tutors were referred as "Pseudo Tutors." However, we have since renamed them more mnemonically as "example-tracing tutors."

explore the use of personalization in a web-based intelligent tutoring setting, as we are doing in our work.

The Clark and Mayer worked example principle proposes that an e-Learning course should present learners with some step-by-step solutions to problems (i.e., *worked examples*) rather than having them try to solve all problems on their own. Interestingly, this principle also runs counter to many people's intuition and even to research that stresses the importance of "learning by doing" [7].

The theory behind worked examples is that solving problems can overload limited working memory, while studying worked examples does not and, in fact, can help build new knowledge [8]. The empirical evidence in support of worked examples is more established and long standing than that of personalization. For instance, in a study of geometry by Paas [9], students who studied 8 worked examples and solved 4 problems worked for less time and scored higher on a posttest than students who solved all 12 problems. In a study in the domain of probability calculation, Renkl [10] found that students who employed more principle-based self-explanations benefited more from worked examples than those who did not. Research has also shown that mixing worked examples and problem solving is beneficial to learning. In a study on LISP programming [11], it was shown that alternating between worked examples and problem solving was more beneficial to learners than observing a group of worked examples followed by solving a group of problems.

Previous ITS research has investigated how worked examples can be used to help students *as* they problem solve [12][13]. Conati's and VanLehn's SE-Coach demonstrated that an ITS can help students self-explain worked examples [14]. However, none of this prior work explicitly studied how worked examples, presented separately from supported problem solving as complementary learning devices, might provide added value to learning with an ITS and avoid cognitive load [8]. Closest to our approach is that of Mathan and Koedinger [15]. They experimented with two different versions of an Excel ITS, one that employed an expert model and one that used an intelligent novice model, complemented by two different types of worked examples, "active" example walkthroughs (examples in which students complete some of the work) and "passive" examples (examples that are just watched). The "active" example walkthroughs led to better learning but only for the students who used the expert model ITS. However, a follow-up study did not replicate these results [16]. This work, as with the other ITS research mentioned above, was not done in the context of a web-based ITS.

2   The Hypotheses and the Stoichiometry Tutor to Test Them

Given the evidence about personalization and worked examples, our intent in the study described in this paper was to explore the following hypotheses:

H1: The combination of personalized language and worked examples, used in conjunction with a supported problem-solving environment (i.e., an ITS), can improve learning in an e-Learning system.

H2: The use of personalized language in a supported problem-solving environment can improve learning in an e-Learning system.

H3: The use of worked examples in a supported problem-solving environment can improve learning in an e-Learning system.

We tested these hypotheses using a Stoichiometry Tutor developed with CTAT. Stoichiometry is the basic math required to solve elementary chemistry problems and is typically learned in the 10th or 11th grade in U.S. high schools. Solving a stoichiometry problem involves understanding basic chemistry concepts, such as the mole, unit conversions, and Avogadro's number, and applying those concepts in solving simple algebraic equations.

Figure 1: The Stoichiometry Intelligent Tutor

The Stoichiometry Tutor and an example of a typical stoichiometry problem (a "personal" version used in the study) are shown in Figure 1. To solve this problem the student must first express the goal, given a problem statement and an initial value. The student has requested a hint, which suggests that they should express the units of the result (i.e., grams or "g," see the highlighted cell, further indicated by the box with an asterisk). After providing the goal, the student must fill in the other terms of the equation to convert the given value to the goal value. Each term, expressed as a ratio, is used to cancel the units and substance of previous terms. The full solution to the problem in Figure 1, with cancelled terms highlighted, is:

$$(0.58 \text{ mol AsO2- } / 100 \text{ kL solution}) * (1 \text{ kL solution } / 1000 \text{ L solution}) * (106.9 \text{ g AsO2- } / 1 \text{ mol AsO2-}) = 0.00062 \text{ g AsO2- } / 1 \text{ L solution}$$

The student is also asked to provide a rationale for each term of the equation (see the "Reason" field below each term in Figure 1). So, for instance, the initial term in the equation of Figure 1 is a "Given Value," since it is provided in the problem statement. Notice that the problem statement and the hint contain first- and second-person pronouns (e.g., "Did you know the WHO recommended limit for arsenic in drinking water is …") instead of more formal, impersonal language (e.g., "Suppose the WHO

4

recommended limit for arsenic in drinking water is…"). The tutor also provides context-specific error messages when the student makes a mistake during problem solving. As with the problem statement and hints, there are personal/impersonal versions of all error messages.

The Stoichiometry Tutor was developed as an example-tracing tutor within CTAT [3]. In particular, after the graphical user interface (GUI) of Figure 1 was created, using the Flash programming language, problem solutions were demonstrated and all hints and error messages were added as "annotations" to the resulting behavior graphs, the knowledge structure that represents individual example-tracing tutor.

3   Description of the Study and Results

To test our hypotheses and the effect of personalized language and worked examples on (e-)learning, we executed a 2 x 2 factorial design, depicted in Figure 2. One independent variable was personalization, with one level being *impersonal* instruction, feedback, and hints and the other *personal* instruction, feedback, and hints. The other independent variable was worked examples, with one level being *supported problem solving only* and the other *supported problem solving and worked examples*. In the former condition, subjects only solve problems; no worked examples are presented. In the latter condition, subjects alternate between observation of a worked example and solving of a problem. As discussed previously, this alternating technique has yielded better learning results in prior research [11].

	Impersonal Instruction, Feedback, and Hints	**Personal Instruction, Feedback, and Hints**
Supported Problem Solving Only	Impersonal / Problem Solving (Condition 1)	Personal / Problem Solving (Condition 2)
Supported Problem Solving and Worked Examples	Impersonal Worked (Condition 3)	Personal / Worked (Condition 4)

Figure 2: The 2 x 2 Factorial Design

If hypothesis H1 is correct, one would expect the subjects in Condition 4 (Personal / Worked) to exhibit significantly larger learning gains than the other conditions, since this is the only condition with both personalized feedback and worked examples. To confirm hypothesis H2, Conditions 2 and 4 should lead to significantly greater learning gains than Conditions 1 and 3 (i.e., a *main effect* for the personalization independent variable). Finally, to confirm hypothesis H3, one would expect Conditions 3 and 4 to exhibit significantly greater learning gains than Conditions 1 and 2 (i.e., a *main effect* for the worked examples independent variable).

The study was executed at the University of British Columbia (UBC) as an optional, online activity in two courses: Intro to Chemistry for majors and Intro to Chemistry for non-majors. Subjects were offered $20 Canadian for completing the study. A total of 1720 students, primarily freshmen and sophomores, were enrolled in the chemistry-for-majors course and 226 were enrolled in the non-majors course. A

total of 240 students started the study, and 69 fully completed it, with "completion" defined as trying at least one step in the final posttest problem. Subjects were randomly assigned to one of the four conditions of Figure 2.

The subjects first watched a video introducing the study and explaining how to use the web-based interface. All subjects were then given an online pretest of nine stoichiometry problems. The pretest (and, later, the posttest) problems were solved using the web-based interface of Figure 1, with feedback and hints disabled. Problems in the pretest (and the posttest) were presented to the subjects in order of difficulty, from relatively easy to fairly difficult. The subjects then worked on 15 "study problems," problems presented according to the different experimental conditions of Figure 2. All of the worked examples in the study were solved using the tutor interface, captured as a video file, and narrated by a chemistry expert. During the solving of the 15 study problems, the subjects were also presented with various instructional videos to instruct them on stoichiometry concepts such as dimensional analysis and molecular weight. After completing the 15 study problems, the subjects were asked to take a posttest of nine problems, analogous in difficulty to the pretest.

All individual steps taken by the students in the stoichiometry interface were logged and scored as correct or incorrect. A score between 0 and 1.0 was calculated for each student's pre and posttest by dividing the number of correct steps by the total number of possibly correct steps. On the pretest there was a possibility of 231 correct steps; on the posttest there was a total of 222.

To test for the effects of worked examples and personalization, a 2 x 2 analysis of variance (ANOVA) was conducted on the difference scores (i.e., post - pre)[2]. For this analysis, there were two between-subjects variables: personalization (personal problems, impersonal problems) and worked examples (worked examples, problem solving). To test for the effect of time (i.e., the overall difference between pre and post across all conditions), a 2 x 2 x 2 repeated measure ANOVA was conducted. For this analysis, the within-subject variable was time (pretest = time 1, posttest = time 2)[3].

Prior to conducting the analyses, six subjects were deleted from the subject pool, four due to sparse pretest and/or posttest responses (indicating they did not seriously attempt to solve the test problems) and two because they were identified as univariate outliers. After data screening, *N* was adjusted from 69 to 63 (Cond. 1 = 16; Cond. 2 = 16; Cond. 3 = 19; Cond. 4 = 12).

The independence assumption underlying the ANOVA was not violated, as each participant was tested in only one of the two personalization conditions and in only one of the two worked problem conditions. A two-way ANOVA of the pretest scores showed no significant differences between the four conditions, indicating that subjects had been successfully randomized. Descriptive statistics indicated that the posttest

---

[2] We also conducted 2x2 ANOVAs using four other common pre/posttest calculations (i.e., (1) (post – pre) / (1.0 – pre), (2) (post-pre) / pre, (3) (post – pre) / ((pre + post) /2), and (4) If post > pre, then (post-pre) / (1.0 – pre) else, (post-pre) / pre [17]). The results were very similar across all analyses, so only the difference score analysis is reported here.

[3] While we could have performed just a 2 x 2 x 2 ANOVA to test for the effects of all of the independent variables, the 2 x 2 ANOVAs provide a much clearer, more intuitive depiction of the effects of worked examples and personalization.

scores were slightly skewed. This departure from normality influences the significance of the ANOVA results, especially with a small sample size ($N= 63$).

Table 1 shows the means and standard deviations of the pretest, posttest, and the difference scores across all conditions. Table 2 shows the result of the 2 x 2 ANOVA on the difference scores ($N = 63$). As shown in Table 2, the main effect of personalization is non-significant ($p= .697$). The impersonal conditions performed slightly better than the personal conditions ($M_{impersonal} = .1455$ vs. $M_{personal} = .1292$). The main effect of worked examples is also non-significant ($p= .828$). The worked example conditions performed slightly better than the supported problem solving conditions ($M_{worked\ examples} = .1401$ vs. $M_{supported\ problem\ solving} = .1365$). The interaction between problem solving and personalization is also non-significant, ($p= .155$).

Table 1: Means & Std Dev. for Pretests, Posttests, and Difference Scores, i.e., (post–pre)

| | Supported Problem Solving | | | | Worked Examples | | | | | |
| | Impersonal (Condition 1) | | Personal (Condition 2) | | Impersonal (Condition 3) | | Personal (Condition 4) | | Total | |
	M	SD	M	SD	M	SD	M	SD	M	SD
Pretest	.7719	.1043	.8231	.1176	.7810	.1554	.7349	.1358	.7806	.1310
Posttest	.9412	.0411	.9268	.0469	.9066	.1160	.8979	.0783	.9189	.0788
Difference Scores	.1692	.1113	.1037	.1185	.1255	.1853	.1630	.1137	.1383	.1392

Table 2: 2 x 2 ANOVA on Difference Scores

Source	df	SS	MS	F	p	η
Between Subjects						
Personalization	1	.003	.003	.153	.697	.003
Worked Examples	1	.001	.001	.048	.828	.001
Personalization x Worked Examples	1	.041	.041	2.074	.155	.034
Error (Between)	59	1.157	.020			

Table 3 shows the results of the 2 x 2 x 2 ANOVA we conducted to assess the effect of time. As can be seen, the effect of time is significant ($p= .000$).

Because of the indeterminate results, we decided to further explore the data. In particular, since we suspected that a large portion of our subject pool was relatively experienced in chemistry, we divided our subjects into three groups: "novices" (18 who scored 71% or below on the pretest), "proficient" (22 who scored greater than 71% but less than or equal to 84%), and "expert" (23 who scored equal to or greater than 84%). The novice group achieved the largest mean score gain from pre to posttest by a wide margin (novice increase in mean: 29%, proficient: 12%, expert: 3%). We also ran a two-way ANOVA on the novice group and, while there was still no significant main effect of worked examples, the difference between the worked examples conditions and

the supported problem solving conditions was greater than that of the population as a whole. Taken together, these findings provide some indication that worked examples may have had more effect on students with less expertise.

Table 3: 2 x 2 x 2 repeated measure ANOVA

Source	df	SS	MS	F	p	η
Between Subjects						
Personalization	1	.001	.001	.046	.830	.001
Worked Examples	1	.039	.039	2.878	.095	.047
Personalization x Worked Examples	1	.016	.016	1.190	.280	.020
Error (Between)	59	.798	.014			
Within Subjects						
Time	1	.604	.604	61.647*	.000	.511
Error (Within)**	59	.578	.010			
Total	125					

* $p < .001$
** Time x Personalization, Time x Worked Examples, Time x Personalization x Worked Examples are not shown because they are not of interest to the analysis of this study

Since we suspected that the lack of a worked examples main effect was due, at least in part, to subjects not fully watching and/or processing the worked examples, we also did an analysis of video viewing. We found that percentage of subjects who fully watched the worked examples started at 48% for the first worked example, but dropped, in almost monotonic fashion, to 26% by the last worked example.

4   Discussion of the Study Results

Our hypotheses that personalization and worked examples would improve learning were not supported, as there were no significant main effects or a significant interaction effect. We were particularly surprised about the lack of effect from worked examples, given the substantial supporting body of research. Why might this have occurred? We have three hypotheses.

First, the subjects in our study may have had too much stoichiometry expertise, potentially offsetting the working memory advantage afforded by studying examples. Previous research has shown that the more expert students are, the less they gain from studying worked examples and the more they gain from problem solving [18]. The students in the UBC chemistry class for majors, which likely constituted a larger proportion of the subjects in our study, would almost surely have had stoichiometry in high school. In addition, because the study was optional and a small percentage of the possible subject pool actually finished the study (< 5%), it is also likely that the majority of the subjects who finished participated precisely because they were confident they already knew the material. This conjecture is supported by the relatively high pretest scores across all conditions and an optional post-study survey in which a high percentage of the subjects claimed the material was not difficult.

Second, worked examples have been shown to be most effective when the learner self-explains them [19] or tackles them as "completion" problems (i.e., the subject

observes part of the solution, then completes the problem on their own, see e.g., [1], pgs. 177-178). Both approaches help the learner process solutions at a deeper level. The worked example videos in our study were only observed; the subjects were not prompted to self explain or complete them. Also, as discussed above, a high percentage of subjects did not fully watch the videos. On the other hand, active processing of worked examples did not definitively lead to better results in the Mathan studies [15]; only the subjects who used the expert model ITS showed significant learning gains by actively processing examples and this result was not replicated in a second study.

Third, and most intriguing with respect to worked examples employed in conjunction with intelligent tutors, it may be that the tutoring received by the subjects simply had much more effect on learning than the worked examples (or personalization). As mentioned above, there was a significant difference between the pretest and posttest across all conditions. Thus, the subjects did appear to learn, but the personalization and worked example interventions did not appear to make the difference. Since we didn't directly compare supported problem solving (i.e., with tutoring) with regular problem solving (i.e., without tutoring), we can't definitively attribute the positive learning effects to tutoring. On the other hand, the analysis we did on the 18 novice performers indicates that the worked examples may have had the desired effect on at least that class of subjects.

Assuming the explanation of these results is not that tutoring "swamped" the effects of the other interventions, we have several hypotheses as to why personalization did not make a difference. First, many of our subjects may have been non-native English speakers and thus missed the nuances of personalized English. We did not collect demographic information, but the UBC chemistry professor said "perhaps more than 50% of the students were non-native English speakers." Second, as with worked examples, the fact that our subjects were not novices may have made it difficult to get an effect. Finally, perhaps our conceptualization and implementation of personalization was not as socially engaging as we had hoped. In a recent study [20], Mayer and colleagues investigated the role of politeness in the conversational style of an on-screen tutor. In a polite version of their system face-saving constructions were used such as, "You could press the ENTER key", and in the direct version, the tutor used direct constructions such as, "Press the ENTER key." Students learned more with the polite tutor, suggesting that providing an on-screen agent with social intelligence makes a difference. In other words, this study suggests that it is not just conversational first and second person language, such as that employed by the stoichiometry tutor, that makes a difference, but the development of a real social relationship with the learner.

5   Conclusions

In this, our first experiment applying Clark and Mayer's e-Learning principles to web-based intelligent tutoring, our results were somewhat disappointing. None of our three hypotheses, relating to the affordances of personalization and worked examples, was supported. On the other hand, we have demonstrated that web-based tutoring *can* be effective, as shown by the significant learning gains across all conditions. In sum-

mary, because of the issues cited above, we view this as a preliminary study whose purpose was to help us develop a workable methodology for testing the effects of personalization and worked examples.

We are currently running a second experiment aimed at U.S. high school students, a pool of subjects that is arguably more appropriate for our stoichiometry materials. We have modified the worked examples, so that subjects must fully watch the videos and correctly answer several self-explanation questions before moving on to the next problem. We conjecture that this will increase the effectiveness of the examples but, as witnessed by the mixed results of the Mathan studies, it is not certain this will make a difference. With respect to personalization, we did not make changes to the study problems, since we believe a more appropriate subject pool of (mostly) native English speakers may lead to different results.

References

1. Clark, R. C. and Mayer, R. E. (2003). *e-Learning and the Science of Instruction*. Jossey-Bass/Pfeiffer.
2. Murray, T., Ainsworth, S., & Blessing, S. (eds.) (2003). *Authoring Tools for Advanced Technology Learning Environments: Toward Cost-Effective, Adaptive, Interactive, and Intelligent Educational Software*. Kluwer Acad. Publishers.
3. Koedinger, K. R., Aleven, V., Heffernan, N., McLaren, B. M., and Hockenberry, M. (2004) Opening the Door to Non-Programmers: Authoring Intelligent Tutor Behavior by Demonstration; *Proceedings of ITS-2004*, 162-174.
4. Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *Int'l Journal of Artificial Intelligence in Education, 8*, 30-43.
5. Beck, I., McKeown, M. G., Sandora, C., Kucan, L., and Worthy, J. (1996). Questioning the author: A year long classroom implementation to engage students in text. *Elementary School Journal, 96*, 385-414.
6. Moreno, R. and Mayer, R. E. (2000). Engaging students in active learning: The case for personalized multimedia messages. *Journal of Ed. Psych., 93*, 724-733.
7. Kolb, D. A. (1984). *Experiential Learning - Experience as the Source of Learning and Development*, Prentice-Hall, New Jersey. 1984.
8. Sweller, J. (1994). Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction, 4*, 295-312.
9. Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Ed. Psych., 84*, 429-434.
10. Renkl, A. (1997). Learning from Worked-Out Examples: A Study on Individual Differences. *Cognitive Science*, 21, 1-29.
11. Trafton, J. G. and Reiser, B. J. (1993). The contributions of studying examples and solving problems to skill acquisition. In M. Polson (Ed.) *Proceedings of the 15th annual conference of the Cognitive Science Society*, 1017-1022.
12. Gott, S. P., Lesgold, A., & Kane, R. S. (1996). Tutoring for Transfer of Technical Competence. In B. G. Wilson (Ed.), *Constructivist Learning Environments*, 33-48, Englewood Cliffs, NJ: Educational Technology Publications.

13. Aleven, V. & Ashley, K. D. (1997). Teaching Case-Based Argumentation Through a Model and Examples: Empirical Evaluation of an Intelligent Learning Environment, *Proceedings of AIED-97*, 87-94.

14. Conati, C. and VanLehn, K. (2000). Toward Computer-Based Support of Meta-Cognitive Skills: a Computational Framework to Coach Self-Explanation. *Int'l Journal of Artificial Intelligence in Education, 11*, 398-415.

15. Mathan, S. and Koedinger, K. R. (2002). An Empirical Assessment of Comprehension Fostering Features in an Intelligent Tutoring System. *Proceedings of ITS-2002*. Lecture Notes in Computer Science, Vol. 2363, 330-343. Berlin: Springer-Verlag

16. Mathan, S. (2003). *Recasting the Feedback Debate: Benefits of Tutoring Error Detection and Correction Skills*. Ph.D. Dissertation, Carnegie Mellon Univ., Pitts., PA.

17. Bonate, P. L. (2000). *Analysis of Pretest-Posttest Designs*. Chapman & Hall/CRC.

18. Kalyuga, S., Chandler, P., Tuovinen, J., and Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Ed. Psych., 93*, 579-588.

19. Chi, M.T.H., Bassok, M., Lewis, M. W., Reimann, P., and Glaser, R. (1989). Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems. *Cognitive Science*, 13, 145-182.

20. Mayer, R. E., Johnson, L., Shaw, E., & Sahiba, S. (2006). Constructing computer-based tutors that are socially sensitive: Politeness in educational software. International Journal of Human Computer Studies, 64, 36-42.