McLaren, B. M., Lim, S., Yaron, D., & Koedinger, K. R. (2007). Can a Polite Intelligent Tutoring System Lead to Improved Learning Outside of the Lab? In the *Proceedings of the 13th International Conference on Artificial Intelligence in Education* (AIED 2007), 433-440.

Can a Polite Intelligent Tutoring System Lead to Improved Learning Outside of the Lab?

Bruce M. MCLAREN, Sung-Joo LIM, David YARON, and Ken KOEDINGER
Carnegie Mellon University
bmclaren@cs.cmu.edu, sungjol@andrew.cmu.edu, yaron@cmu.edu,
koedinger@cs.cmu.edu

Abstract. In this work we are investigating the learning benefits of e-Learning principles (a) within the context of a web-based intelligent tutor and (b) in the "wild," that is, in real classroom (or homework) usage, outside of a controlled laboratory. In the study described in this paper, we focus on the benefits of politeness, as originally formulated by Brown and Levinson and more recently studied by Mayer and colleagues. We test the learning benefits of a stoichiometry tutor that provides polite problem statements, hints, and error messages as compared to one that provides more direct feedback. Although we find a small, but not significant, trend toward the polite tutor leading to better learning gains, our findings do not replicate that of Wang *et al*., who found significant learning gains through polite tutor feedback. While we hypothesize that an e-Learning principle such as politeness may not be robust enough to survive the transition from the lab to the "wild," we will continue to experiment with the polite stoichiometry tutor.

1. Introduction

The plethora of e-Learning materials available on the web today raises an important question: Does this new technology, delivered to students using new ways of presenting text, graphics, audio, and movies, lead to improved learning? It is important not only to provide easy access to learning technology but also to investigate scientifically whether that technology makes a difference to learning. Mayer and other educational technology researchers have proposed and investigated a variety of e-Learning principles, such as personalization (using conversational language, including first and second-person pronouns, in problem statements and feedback), worked examples (replacing some practice problems with worked examples), and contiguity (placing text near the graphics or pictures it describes), and have run a variety of (predominantly) lab-based experiments to test their efficacy [1]. The evidence-based approach taken by these researchers is important learning science research.

We are also interested in a systematic investigation of e-Learning principles. However, we wish to explore two aspects of e-Learning principles that have been left largely unexplored by previous researchers. First, our goal is to test the principles in the context of a web-based intelligent tutoring system (ITS), rather than in a standard e-Learning environment, which typically provides less feedback for and support of learners. Our main interest is in understanding whether and how the principles can supplement and extend the learning benefits of a tutoring system that runs on the web. Second, as a project within the Pittsburgh Science of Learning Center (PSLC) (www.learnlab.org), we wish to explore one of the key concepts underpinning the PSLC's theoretical framework and approach: the testing of learning interventions in the "wild" (i.e., in a live classroom or homework setting) rather than in a tightly controlled lab setting. The PSLC espouses an approach in which *in vivo* studies are the primary mechanism for experimentation; similar to the way studies are done in medical research

hospitals (http://www.learnlab.org/clusters/PSLC_Theory_Frame_June_15_2006.pdf).

We have initially focused our attention on two of the e-Learning principles mentioned above, in particular, personalization and worked examples. In two *in vivo* studies, in which we applied these principles in the context of an intelligent, web-based tutor, we were not able to replicate the learning benefits found by other researchers in more controlled, lab-based settings. Using a 2 x 2 factorial design, we found that personalization and worked examples had no significant effects on learning. The first study was conducted with 69 university students [2]. The second study, the results of which have not been published, was conducted with 76 students at two suburban U.S. high schools. In both of these studies there was a significant learning gain between the pre- and posttest across all conditions, in line with well-established findings from previous studies of the benefits of intelligent tutoring (e.g., [3, 4]).

So why didn't we achieve learning gains when adding these two e-Learning principles to a tutoring system? One possibility is that the tutoring may simply have had much more effect on learning than either the worked examples or personalization. The students learned at a significant level in both studies but much of that learning may have been induced by the support of the tutor. It is also possible that the principles did not survive the transition from the lab to *in vivo* experimentation. Most of the worked example experiments and all of the personalization experiments cited by Clark and Mayer [1] were conducted in tightly controlled lab environments of relatively short duration (< 1 hour), while our experiments were conducted in messier, real-world settings (i.e., the classroom or at home over the Internet) in which students used the tutor for 3 to 5 hours. It may also be that the principles only work for certain domains or for certain populations. For instance, some research has demonstrated that novices tend to benefit more from worked examples than more advanced students [5].

Another possible explanation, one that focuses on the personalization principle and is the main topic of the current paper, is that the personalization principle may need refinement and/or extension. In particular, perhaps our conceptualization and implementation of "personalization" was not as socially engaging and motivating as we had hoped. Mayer suggested this to us, after reviewing some of the stoichiometry tutor materials (personal communication) and based upon recent research he and colleagues have done in the area of *politeness* [6, 7, 8]. While our stoichiometry tutor is faithful to the personalization principle [1], it may be that the conversational style proposed by the principle is simply not enough to engage learners. Thus, we decided to "upgrade" the principle in its application to the stoichiometry tutor by including politeness. We did this by modifying all of the problem statements, hints, and error messages to create a new "polite" version of the stoichiometry tutor. We then executed a new *in vivo* experiment to test whether the new, polite version of the tutor led to greater learning gains than a more direct version.

In this paper, we briefly review the research on personalization and politeness in e-Learning, describe how we've created a polite version of the stoichiometry tutor, present and discuss the experiment we ran that tested whether the polite tutor improved learning, and conclude with hypotheses about our finding and proposed next steps.

2. Research on Personalization and Politeness in e-Learning

The personalization principle proposes that informal speech or text is more supportive of learning than formal speech or text in an e-Learning environment. E-Learning research in support of the personalization principle has been conducted primarily by Mayer and colleagues, with a total of 10 out of 10 studies demonstrating

deeper learning with personalized versions of e-Learning material, with a strong median effect size of 1.3 [9]. All of these studies focused on the learning of scientific concepts in e-Learning simulation environments and compared a group that received instruction in conversational style (i.e., the personalized group) with a group that received instruction in formal style (i.e., the nonpersonalized group). Note, however, that all of these studies were tightly controlled lab experiments with interventions of very short duration, some as short as 60 seconds, and none were conducted in conjunction with an intelligent tutoring system.

Based on the work of Brown and Levinson [10], Mayer and colleagues have more recently performed a series of studies to investigate whether "politeness" in educational software, in the form of positive and negative face saving feedback, can better support learners. They have implemented positive and negative face feedback in the context of the Virtual Factory Teaching System (VFTS), a factory modeling and simulation tutor. Positive face refers to people's desire to be accepted, respected, and valued by a partner in conversation, while negative face refers to the desire of people not to be controlled or impeded in conversation. In a *polite* version of VFTS they have developed, constructions such as, "You could press the ENTER key" and "Let's click the ENTER button" were used. Such statements are arguably good for positive face, as they are likely to be perceived as cooperative and suggestive of a common goal, as well as for negative face, as they are also likely to be perceived as respectful of the student's right to make his or her own decisions. In the *direct* version of VFTS, the tutor used more imperative, direct feedback such as, "Press the ENTER key" and "The system is asking you to click the ENTER button." These statements are arguably not supportive of positive face, as they do not suggest cooperation, or of negative face, as they are likely to be perceived as limiting the student's freedom.

In a preliminary study run by Mayer *et al*. [7] students were asked to evaluate the feedback of the tutor. The results indicated that learners are sensitive to politeness in tutorial feedback, and that learners with less computer experience react to the level of politeness in language more than experienced computer users. Wang *et al*. [6] ran students through a Wizard-Of-Oz study, with some students using the polite tutor and some using the direct tutor, that showed students liked working with the polite tutor more than the direct tutor and did slightly, but not significantly, better in learning outcome when using the polite tutor. Finally, in the study run by Wang *et al*. [8] in which 37 students were randomly assigned either to a polite tutor group or to a direct tutor group. The students who used the polite tutor scored significantly higher on a posttest. In sum, these studies suggest that it is not just first and second person conversational feedback, such as that proposed by the personalization principle and employed by the stoichiometry tutor, that makes a difference in motivating students and promoting better learning, but instead the level of politeness in that feedback. The Wang *et al*. study also showed that this effect could be achieved in the context of a tutor, something we are also interested in.

Not all research supports the idea that politeness will benefit learning and tutoring. For instance, Person *et al*. studied human tutoring dialogues and suggest that politeness could, under some circumstances and in different domains, inhibit effective tutoring [11]. They also relied on Brown and Levinson as a framework of investigation and found that different steps in the tutoring process appear to be more or less likely to benefit from politeness. However, these findings were not subjected to empirical study.

3. The Stoichiometry Tutor

The Stoichiometry Tutor, developed using the Cognitive Tutor Authoring Tools [12], and an example of a typical stoichiometry problem (a "polite" version) are shown in Figure 1. Solving a stoichiometry problem involves understanding basic chemistry concepts, such as the mole and unit conversions, and applying those concepts in solving simple algebraic equations. To solve problems with the tutor the student must fill in the terms of an equation, cancel numerators and denominators appropriately, provide reasons for each term of the equation (e.g., "Unit Conversion'), and calculate and fill in a final result. The tutor can provide student requested-hints, as is shown in Figure 1 (the hint refers to the highlighted cell in the figure), and also provides context-specific error messages when the student makes a mistake during problem solving [2].

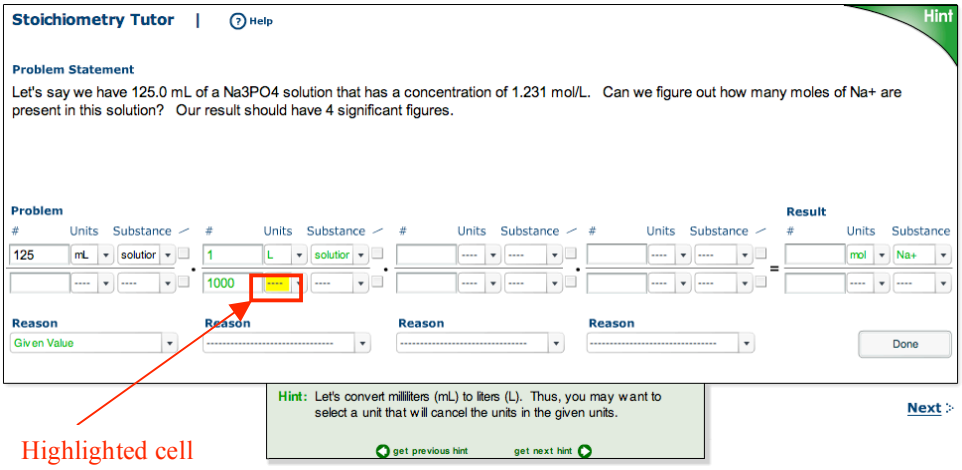


Figure 1: The Stoichiometry Intelligent Tutor

4. Changing the Stoichiometry Tutor From Personalization to Politeness

Given the learning outcomes achieved by Mayer and colleagues we decided to investigate whether altering our stoichiometry tutor's language feedback, making it more polite and improving its positive and negative face, would lead to better learning results. To implement and test this we created a "polite" version of the stoichiometry tutor by altering all of the problem statements, hints, and error feedback of the personalized version of the stoichiometry tutor to make them more polite, following the approach of Mayer *et al*. [7]. In addition, we added "success messages" to some correct responses made by students (in the personalized and direct versions of the tutor, there were no success messages), another positive polite strategy suggested in Wang *et al*. [6]. Examples of the changes we made to create a polite stoichiometry tutor, as well as how these changes compare to the direct and personal versions, are shown in Table 1.

Notice that all of the messages from the polite stoichiometry tutor are intended to be good for both positive and negative face, in comparison to the direct version of the stoichiometry tutor. For instance, the polite problem statement "Can we calculate the number of grams of iron (Fe) that are present in a gram of hematite (Fe2O3)?" provides positive politeness by suggesting cooperation and a common goal between the student and the tutor (use of "we") and negative politeness through giving the student freedom of choice (use of "Can" and phrasing the problem as a question; using "should" in the second sentence of the problem statement). In contrast, the analogous problem

statement in the direct stoichiometry tutor ("How many grams of iron (Fe) are present in a gram of hematite (Fe2O3)?") is lacking in both positive politeness (i.e., no sense of collaboration suggested) and negative politeness (i.e., the student's freedom of action is not acknowledged, as the wording of the statement assumes the student *will* do the problem). For similar reasons, all of the other problem statements, hints, and error feedback examples taken from the polite stoichiometry tutor in Table 1 have arguably more positive and negative politeness than the corresponding statements in the direct stoichiometry tutor. In addition, the polite version of the tutor is the only one to provide positive politeness in the form of success messages.

Table 1: Examples of Language Diffs. Between the Polite, Direct, and Personal Versions of the Stoich Tutor

	Polite Stoichiometry Tutor	**Direct Stoichiometry Tutor ("Impersonal" version, see [2])**	**Personal Stoichiometry Tutor ("Personal" version, see [2])**
Problem Stmts.	Can we calculate the number of grams of iron (Fe) that are present in a gram of hematite (Fe2O3)? Our result should have 5 significant figures.	How many grams of iron (Fe) are present in a gram of hematite (Fe2O3)? The result should have 5 significant figures.	Can you calculate and tell me how many grams of iron are present in a gram of hematite? Your result should have 5 significant figures.
Hints	Let's calculate the result now.	The goal here is to calculate the result.	You need to calculate the result.
	Do you want to put 1 mole in the numerator?	Put 1 mole in the numerator.	You need to put 1 mole in the numerator.
Error Msgs.	You could work on a composition stoichiometric relationship in this term. Are grams part of this relationship?	This problem involves a composition stoichiometric relationship in this term. Grams are not part of this relationship.	You need to work on a composition stoichiometric relationship in this term. Grams are not part of this relationship.
	Won't we need these units in the solution? Let's not cancel them, Ok?	No, these units are part of the solution and should not be cancelled.	Won't you need these units in the solution? If so you shouldn't cancel them.
Success Msgs.	Super job, keep it up.	*None*	*None*

5. Method

We conducted a study using the 2 x 2 factorial design discussed previously and shown in Table 2. One independent variable was politeness, with one level *polite* problem statements, feedback, and hints (i.e., use of the polite stoichiometry tutor) and the other level *direct* instruction, feedback, and hints (i.e., use of the direct stoichiometry tutor). The other independent variable was worked examples, with one level being *tutored only* and the other level *tutored and worked examples*. For the former level, subjects only solve problems with the tutor; no worked examples are presented. In the latter, subjects alternate between observation and (prompted) self-explanation of a worked example and solving of a problem with the aid of the tutor (either the polite or direct tutor, dependent on the condition). Note that although our main interest in this study was the effect of politeness, we decided to keep the design of the McLaren *et al.* experiment [2], since we were curious whether the null result with respect to worked examples in the previous experiment would hold again.

The study was conducted with 33 high school students at a suburban high school in Pittsburgh, PA. The materials were presented as an optional extra credit assignment in a college prep chemistry class, and the students could either do the work during a school

study hall or at home. Subjects were randomly assigned to one of the four conditions in Table 2.

Table 2: The 2 x 2 Factorial Design

	Polite Instruction	**Direct Instruction**
Tutored Only	Polite / Tutored (Condition 1)	Direct / Tutored (Condition 2)
Tutored and Worked Examples	Polite / Worked (Condition 3)	Direct / Worked (Condition 4)

All subjects were first given an online *pre-questionnaire* with multiple-choice questions designed to measure confidence (e.g., "I have a good knowledge of stoichiometry" Not at all, … Very Much) and computer experience (e.g., "How many hours a week do you normally use a computer?" < 1 hr, 1-5 hours, … > 20 hours). All subjects were then given an online pretest of 5 stoichiometry problems. The subjects took the pretest (and, later, the posttest) using the web-based interface of Figure 1, with feedback and hints disabled. The subjects then worked on 10 "study problems," presented according to the different experimental conditions of Table 2. All of the worked examples in the study were solved using the tutor interface, captured as a video file, and narrated by the first author (McLaren). During the solving of the 10 study problems, the subjects were also periodically presented with various instructional videos. After completing the 10 study problems, all subjects were given an online *post-questionnaire* with multiple-choice questions designed to assess their feelings about their experience with the tutor (e.g., "The tutor was friendly to me" Not at all … Almost Always). Finally, the subjects were asked to take a posttest of 5 problems, with problems isomorphic to the pretest.

All individual steps taken by the students in the stoichiometry interface of the pretest and posttest were logged and automatically marked as correct or incorrect. A score between 0 and 1.0 was calculated for each student's pretest and posttest by normalizing the number of correct steps divided by the total number of possibly correct steps.

6. Results

We did a two-way ANOVA (2 x 2) on the difference scores (posttest – pretest) of all subjects. There were no significant main effects of the politeness variable ($F(1, 29) = .66$, $p = .42$) or the worked examples variable ($F(1, 29) = .26$, $p = .61$). To test the overall effect of time (pretest to posttest), a repeated measure ANOVA (2 x 2 x 2) was conducted. Here there was a significant effect ($F(1, 29) = 48.04$, $p < .001$). In other words, students overall – regardless of condition – improved significantly from pretest to posttest. A summary of the results, in the form of a means of adjusted posttest, is shown in Figure 2. These results were very similar to the results of

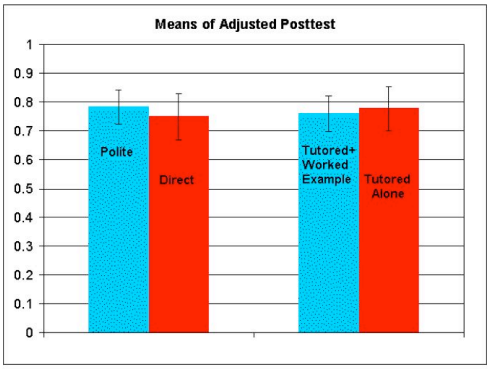


Figure 2: Means of Adjusted Posttest N=33.

McLaren et al. [2], except that in the current study the polite condition did somewhat better than the direct condition, as illustrated in Figure 2, whereas in the previous study the impersonal condition did slightly, but not significantly, better than the personal condition. While the difference is not significant, it nevertheless can be considered a small trend. We did a one-way ANOVA on posttest covariate with pretest that showed a small, but clearly larger, slope from the pretest to posttest for the polite condition.

Since politeness did not seem to support learning, contrary to the findings of Wang *et al*. [8], we did some analysis of the pre- and post-questionnaire responses, similar to Mayer *et al.* [7], to see if students at least noticed and reacted to the positive and negative politeness of the polite stoichiometry tutor's feedback. Students provided answers to the multiple-choice questions shown in the box below, selecting between 5 responses ("Not likely," "Not much," "Somewhat," "Quite a bit," and "Almost Always").

We scaled the responses from 1 to 5 and performed an independent-samples t-test between the polite and direct conditions. Only highlighted question 'g' led to a statistically significant difference between the polite and direct conditions (p=.000), with the polite condition having a higher rating than the direct condition. In other words, the polite stoichiometry tutor's feedback only appeared to be received as especially polite and helpful in how it praised students for doing something right.

a. I liked working with the stoichiometry tutor
b. The tutor helped me identify my mistakes
c. The tutor worked with me toward a common goal
d. The tutor was friendly to me
e. The tutor let me make my own choices
f. The tutor made me follow instructions
g. The tutor praised me when I did something right
h. The tutor was critical of me
i. My relationship with the tutor was improving over time

While we did not find a strong overall effect of subjects finding the polite tutor more polite than the direct tutor, we did find certain groups of students who were sensitive to the more polite statements. For instance, within a "low-confidence group" (i.e., N = 9; subjects who had a score <=5 when summing 3 confidence questions, highest possible score = 15), we found a stronger sensitivity to the polite tutor's statements. The low-confidence students gave higher ratings, at a marginally significant level, to the positive politeness questions e ("The tutor let me make my own choices," p=.09) and c ("The tutor worked with me toward a common goal," p=.119) and lower ratings, again at a marginally significant level, to the negative politeness question f ("The tutor made me follow instructions," p=.112). Within a "low-skill group" (i.e., N = 8; subjects who scored less than 0.5 (50%) on the pretest), we also found positive sensitivity to the polite tutor's statements. The low-skill students gave higher ratings, at a significant level, to positive politeness question i ("My relationship with the tutor was improving over time," p=.041) and higher ratings at a significance level that, while not marginally significant, were quite a bit lower than all others to the positive politeness questions a ("I liked working with the stoichiometry tutor," p=.165) and b ("The tutor helped me identify my mistakes," p=.165).

7. Discussion and Conclusions

Although the polite stoichiometry tutor showed a small trend toward better learning gains than a more direct version of the tutor, our findings, at least to this point, do not support the findings of Wang *et al*. [8] that politeness makes a difference in learning.

There are a variety of possible reasons for this, but our primary hypothesis is that politeness, while making a difference in controlled lab studies of short duration is less

likely to have an effect in a real world study such as ours. One might argue that such a study is subject to methodological problems, such as excessive time on task, etc., but our objective, as part of a learning science center with a particular mission, is to test interventions in just these kinds of circumstances. After all, if educational interventions cannot endure the transition from the lab to the real world, than their usefulness is questionable. As demonstrated in our two studies thus far, the use of an intelligent tutor *did* make a significant difference to learning in a classroom / homework situation; it is the additional e-Learning principles that did not appear to add to learning gains.

Another possibility is that our polite tutor didn't exhibit enough positive and negative politeness to have a real impact on students. There is some evidence of that in our analysis of the post-questionnaire. While there is some evidence that students who used the polite tutor really found it more polite, especially within certain groups such as low-confidence and low-skill students, the overall impression of politeness it left on students does not appear to rise to the level reported in [7]. Our earlier hypothesis that both of the e-Learning principles were "swamped" by the effect of tutoring is potentially refuted by Wang *et al* [8]. In their study, with an N very similar to ours, the effect of tutoring did not overwhelm the effect of politeness. That is, in their study, politeness made a difference even in the midst of tutoring.

In conclusion, however, both our N of 33 and theirs of 37 are relatively small; a power analysis we performed indicated the need for an N = ~320 to reach a power of .8. In an attempt to get a larger effect size and more fully investigate the polite version of the stoichiometry tutors, we are planning a second phase of this study at two suburban New Jersey high schools in early 2007.

Acknowledgements. The PSLC, NSF Grant # 0354420, has provided support for this research.

References

[1] Clark, R. C. & Mayer, R. E. (2003). *e-Learning and the Science of Instruction*. Jossey-Bass/Pfeiffer.

[2] McLaren, B. M., Lim, S., Gagnon, F., Yaron, D., & Koedinger, K. R. (2006). Studying the Effects of Personalized Language and Worked Examples in the Context of a Web-Based Intelligent Tutor; In the *Proceedings of the 8th International Conference on Intelligent Tutoring Systems*, 318-328.

[3] Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, 8, 30-43.

[4] VanLehn, K., Lynch, C., Schulze, K., Shapiro, J., & Shelby. (2005). The Andes Physics Tutoring System: Five Years of Evaluations. In the *Proceedings of AIED-05*. 678-685.

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

[6] Wang, N., Johnson, W. L., Rizzo, P., Shaw, E., & Mayer, R. E. (2005). Experimental Evaluation of Polite Interaction Tactics for Pedagogical Agents. In the *Proceedings of the 2005 International Conference on Intelligent User Interfaces,* 12-19. ACM Press: New York.

[7] Mayer, R. E., Johnson, W. L., Shaw, E. & Sandhu, S. (2006). Constructing Computer-Based Tutors that are Socially Sensitive: Politeness in Educational Software. *Int. J. Human-Computer Studies,* 64, 36-42.

[8] Wang, N., Johnson, W. L., Mayer, R. E., Rizzo, P., Shaw, E., & Collins, H. (in press). The Politeness Effect: Pedagogical Agents and Learning Outcomes. To be published in the *International Journal of Human-Computer Studies*.

[9] Mayer, R. (2005). Principles of Multimedia Learning Based on Social Cues: Personalization, Voice, and Image Principles. *The Cambridge Handbook of Multimedia Learning*, R. E. Mayer (Ed). Chapter 14, 201-212.

[10] Brown, P. & Levinson, S. C. (1987). *Politeness: Some Universals in Language Use*. Cambridge University Press, New York.

[11] Person, N. K., Kreuz, R. J., Zwaan, R. A., & Graesser, A. C. (1995). Pragmatics and Pedagogy: Conversational Rules and Politeness Strategies May Inhibit Effective Tutoring. *Cognition and Instruction* 13, 161-188.

[12] Koedinger, K. R., Aleven, V., Heffernan, N., McLaren, B. M., & Hockenberry, M. (2004) Opening the Door to Non-Programmers: Authoring Intelligent Tutor Behavior by Demonstration; *Proceedings of ITS-2004*, 162-174.