Shallow learning as a pathway for successful learning both for tutors and tutees

Noboru Matsuda (mazda@cs.cmu.edu), **Evelyn Yarzebinski** (eey2@cs.cmu.edu),
Victoria Keiser (keiser@cs.cmu.edu), **Rohan Raizada** (rohan@cs.cmu.edu),
William W. Cohen (wcohen@cs.cmu.edu)
School of Computer Science, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213, USA

Gabriel Stylianides (gabriel.stylianides@education.ox.ac.uk)
Department of Education, University of Oxford
15 Norham Gardens, Oxford, OX2 6PY, UK

Kenneth R. Koedinger (krk@cs.cmu.edu)
School of Computer Science, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213, USA

Abstract

SimStudent is a computational model of learning with its cognitive fidelity of learning being demonstrated especially in the way it makes human-like errors. Using SimStudent as a teachable agent in an interactive peer-learning environment, we have investigated how tutee (i.e., SimStudent) learning affected tutor (i.e., human student) learning. In this paper, we are particularly interested in how tutees' shallow learning affects tutor learning. We are also interested in how the errors that the tutee makes affect tutor learning. The results show that teaching SimStudent on a fixed set of problems makes students easy to tutor SimStudent, which in turn helps students learn, but is likely to allow SimStudent to commit shallow learning, which is harmful for tutor learning. It is thus crucial to let the student detect SimStudent's shallow learning and extend teaching until SimStudent and the student achieve satisfactory competence.

Keywords: Learning by teaching; teachable agent; SimStudent; shallow learning; learning from errors.

Introduction

Studying the effect of learning by teaching through the use of teachable-agent technology is a rapidly growing research field. There have been a number of teachable agents used in empirical classroom studies, for example, Betty's Brain (Biswas, Leelawong, Schwartz, Vye, & Vanderbilt, 2005) and TAAG (Pareto, Arvemo, Dahl, Haake, & Gulz, 2011).

Researchers have explored different aspects of the effect of tutor learning, including learning meta-cognitive skills for self-regulated learning (Biswas, Jeong, Kinnebrew, Sulcer, & Roscoe, 2010), the protégé effect (Chase, Chin, Oppezzo, & Schwartz, 2009), the adaptive assistance (Walker, Rummel, & Koedinger, 2009), and the effect of self-explanation (Matsuda, Keiser, et al., 2012).

The teachable agent we have developed is called SimStudent. SimStudent is a machine-learning agent that learns procedural problem-solving skills from examples (Matsuda, Cohen, Sewall, Lacerda, & Koedinger, 2008). SimStudent can be interactively tutored (aka, learning from tutored problem-solving), and has been integrated into an on-line, game-like learning environment, called APLUS (Artificial Peer Learning environment Using SimStudent). The current version of APLUS allows students to learn Algebra equations by teaching SimStudent. Using APLUS, we have conducted a number of classroom studies to advance cognitive and social theories of tutor learning (Matsuda, Keiser, et al., 2012; Matsuda et al., 2011).

The goal of this paper is to investigate the relationship between tutee- and tutor-learning. As previous empirical studies show (e.g., Cohen, 1994), peer tutoring is known to be beneficial both for tutors and tutees. We thus hypothesize that there must be a strong correlation between SimStudent's and human students' learning. We are particularly interested in how a tutee's shallow learning affects tutor learning. When tutoring, the tutor might fail to detect the tutee's shallow learning by observing the tutee's satisfactory performance at the surface level without actually probing for underlying deep understanding of the domain knowledge. However, if there is actually a symbiotic relationship between tutee and tutor learning, then the tutee's shallow learning should be detrimental to tutor learning.

We are also interested in studying how tutee errors help not only tutee but also tutor learning. In a previous experiment, we studied a theoretical account of the impact of corrective feedback on SimStudent's learning (Matsuda, et al., 2008). We found that committing errors and receiving explicit corrective feedback facilitates tutee learning. On the other hand, it is also known that (human) students learn by explaining erroneous worked-out examples (Grosse & Renkl, 2006; Siegler, 2002). Therefore, tutee errors would also help tutors learn when tutors explain errors committed by tutees. The cognitive fidelity of SimStudent has been demonstrated especially in the way it makes human-like induction errors to learn incorrect skills and hence makes human-like errors when solving problems (Matsuda, Lee, Cohen, & Koedinger, 2009). Therefore, using SimStudent to understand how tutee errors affect tutor learning would be a valid research methodology.

To test the above hypotheses, we conducted a secondary data analysis using the data we collected from our previous classroom studies in which we tested the effect of APLUS.

In the remaining sections, we will first briefly introduce an overview of SimStudent and APLUS in enough detail to understand the research questions and hypotheses. We then describe the data we analyze and the classroom studies from which the data were collected. Finally, we show results followed by a discussion.

Learning by Teaching SimStudent

SimStudent

SimStudent is a machine-learning agent that learns procedural skills from examples. When serving as a teachable agent, SimStudent commits to guided problem solving. That is, SimStudent attempts to solve problems given by the student, suggesting one step at a time by applying a learned production. SimStudent asks the student about the correctness of the suggestions. If the student provides negative feedback, SimStudent may attempt to provide an alternative suggestion. When SimStudent has no suggestion that receives positive feedback from the student, then it asks the human student to demonstrate the step as a hint.

The student's feedback and hints become examples that SimStudent generalizes using domain specific background knowledge. As a result, SimStudent generates hypotheses about how to solve problems in the form of *production rules*. SimStudent uses a hybrid learning algorithm that involves (1) inductive logic programming to learn when to apply a production rule, (2) a version space to learn upon what to focus attention, and (3) an iterative-deepening depth-first search to learn how to change the problem state.

SimStudent occasionally prompts students to explain their tutoring actions by asking "why" questions. Such questions include (1) the reason for selecting a particular problem to solve, (2) the reason for an incorrect suggestion, and (3) the reason for the student's demonstration.

APLUS

APLUS has a Tutoring Interface on which the student and SimStudent collaboratively solve problems. To pose a problem, the student enters an equation into the first row of the Tutoring Interface. As SimStudent makes suggestions for each step, they are placed into the Tutoring Interface. When SimStudent requires a hint, the student demonstrates the next correct step in the Tutoring Interface.

In the regular version of APLUS, the goal of the student is to tutor SimStudent well enough to pass the Quiz. At any time while tutoring, the student may click on the [Quiz] button. SimStudent's productions learned thus far are applied to a set of Quiz problems, and the summary of the results appears in a separate window showing the correctness of the steps suggested by SimStudent. See (Matsuda, et al., 2011) for more details about APLUS.

There have been two versions of APLUS implemented so far, and each version was used in different classroom studies (see Section "Classroom Studies" for details about the classroom studies). The two versions differ in the structure of the Quiz. In the earlier version, the Quiz problems were fixed. SimStudent took the exact same set of Quiz problems each time it was quizzed. In the later version, the Quiz problems were randomly generated based on a fixed problem *type*. That is, the coefficients and constants were randomly generated each time SimStudent was quizzed.

How does SimStudent commit Shallow Learning?

In APLUS, one potential pit-fall that may induce SimStudent's shallow learning is the usage and structure of the Quiz. In the earlier study (called the Self-Explanation Study), since the problems in the Quiz were fixed, students could have focused on tutoring only those fixed problems. SimStudent's learning might have been "shallow," or overly specific to solve only those problems, which could have also led human students to "shallow" learning.

On the other hand, the problems in the Quiz for the later study (called the Game Show Study) were randomly generated each time SimStudent took a Quiz (although, they are always in the fixed *type*). Therefore, if SimStudent passes the Quiz in Game Show study, it is likely that SimStudent has learned a high quality set of productions – i.e., "deep" learning. In fact, there were 19 SimStudents that passed the Quiz in the Self-Explanation study, but no SimStudents passed the Quiz in the Game Show study.

An example would help to understand SimStudent's shallow learning. In one instance, SimStudent in the Self-Explanation Study learned to divide both sides of the equation in the form of Ax=B, where x is a variable, A is a coefficient, and B is a constant term. The production for division says "divide both sides by a chunk of digits before the variable." The "chunk of digits" by definition only perceives a number before the variable without a sign. This piece of background knowledge was designed to model human student's common induction errors (Matsuda, et al., 2009).

As a consequence, this SimStudent could solve equations Ax=B only when the coefficient A is a positive number. In the fixed set of the Quiz, this SimStudent learned to solve the equations in such a way that it always happened to have a positive coefficient on the last step, i.e., Ax=B (or A=Bx). However, even when the same productions were applied, the randomized Quiz problems sometimes produced negative coefficients when combining like terms or balancing the equation. Because of such an accidental transformation, this SimStudent was not able to pass the randomized quiz.

Research Questions and Hypotheses

This paper addresses the following three research questions and hypotheses.

Q1: *How do tutee and tutor learning relate?* We first hypothesize that SimStudent's learning and human students' learning are correlated. To test this hypothesis, we quantify SimStudent's learning as the "quality" of productions learned by SimStudent. Human students' learning will be quantified using test scores.

Q2: *Is a tutee's shallow learning detrimental to tutor learning?* We hypothesize that letting SimStudents do shallow learning is harmful for tutor learning. To test this hypothesis, we will validate the production rules of SimStudents who passed the fixed Quiz to see if they can also pass the randomly generated Quiz. We will then examine if students who allowed their SimStudents to commit to shallow learning showed poor learning.

Q3: *How do tutee errors influence tutor learning?* We hypothesize that the effect of learning from erroneous examples would apply for tutor learning, that is, detecting SimStudent's errors correctly and explaining those errors would facilitate tutor learning.

Method

Sample

The analysis was done on the data we previously collected from two classroom "*in-vivo*" studies; the Self-Explanation Study (Matsuda, Keiser, et al., 2012) and the Game Show Study (Matsuda, Yarzebinski, et al., 2012). Both sets of data are available (upon request) on the large-scale educational database, DataShop (Koedinger et al., 2010), maintained by Pittsburgh Science of Learning Center.

The data include the outcome data and process data. The outcome data are test scores. Students took pre- and post-tests before and after tutoring SimStudent. The test consists of (1) ten equation-solving items, (2) twelve items to determine if a given operation is a logical next step for a given equation, and (3) five items to identify the incorrect step in a given incorrect solution. The pre- and post-tests are isomorphic.

The process data shows the interaction between individual students and SimStudent. It contains (among other things) problems tutored, feedback provided by the students (and their correctness), steps performed both by students and SimStudent (and their correctness), hints requested by SimStudent, and quiz attempts (and their results).

Classroom Studies

Two classroom studies were conducted in the same school near Pittsburgh, PA, but for different Algebra I classes. The Self-Explanation (SE) study included 111 students from advanced 8th grade and regular & remedial 9th grade classes. The Game Show (GS) study included 141 students from advanced 7th and regular 8th grade classes.

Both studies were conducted as randomized control trials. There were three intervention days (a single class period per day) when students used APLUS. All students took pre- and post-tests before and after the intervention.

For the current study, only the data from the treatment condition in the SE Study and the control condition in the GS Study were used, because the students in those conditions used the same version of SimStudent and APLUS with the same goal for tutoring (i.e., passing the Quiz). As a result, there were 44 students in each condition who took both pre- and post-test and completed the intervention (meaning, they either attended all three days of the intervention or passed the Quiz sooner).

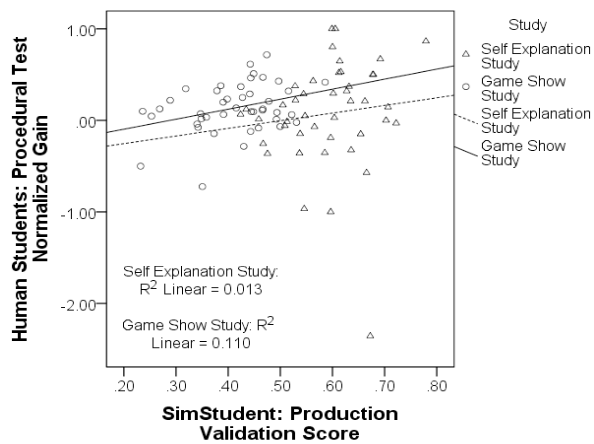

Figure 1: Scatter plot showing SimStudent's validation scores (X-axis) and students normalized gain (Y-axis).

Measures

In the following analysis, human students' learning is measured as the normalized gain of the test score, which is computed as (post-test – pre-test)/(1 – pre-test).

SimStudent's learning is measured as the accuracy of productions learned to solve equations. The productions were applied to a total of 30 equation problems taken from the actual tests that the human students took.[1] For each step in solving an equation, the correctness of each of the applicable productions (i.e., the conflict set) was judged by using the expert model of the Algebra Cognitive Tutor. The *step score* was then calculated as the ratio of correct production firing to all applicable productions. The step score is zero when there are no applicable productions. The *problem score* was computed by averaging the step scores across all steps. Finally, the *validation score* was computed as the average problem score for the 30 equation problems.

The quality of students' responses to SimStudent's "why" questions was also evaluated by three human coders. The coders categorized the student's responses into "deep" and "shallow" responses.

Results

Tutor-tutee Learning Correlation

How does tutee learning correlate with tutor learning? To answer this question, we first tested the correlation between SimStudent's learning gain and human students' learning gain.

Figure 1 shows the scatter plot with SimStudent's learning represented by the validation score (X-axis) and the human students' learning as the normalized gain on the test scores (Y-axis). Data points from the Self-Explanation

[1] In the classroom studies, there was also a delayed-test. Thus, the students took three tests each containing 10 equation problems.

Study and the Game Show Study are represented using circles and triangles, respectively.

There is a significant correlation between SimStudent's learning and human students' learning for the Game Show Study; $r(43)=0.331$, $p<0.05$.

The correlation between SimStudent's learning and human student's learning was not significant for the Self-Explanation Study; $r(43)=0.115$, $p=0.463$. This might be partly because of the large variance in the human student's learning; $M=0.07$, $SD=0.59$.

Was there any difference in SimStudent's and human students' learning between the two studies? The study (SE vs. GS) is a main effect for the SimStudent's validation score; $t(86)=10.488$, $p<0.000$. SimStudents in the Self Explanation Study learn better than the Game Show; mean validation score $M_{SE}=0.59$ ($SD=0.08$) vs. $M_{GS}=0.41$ ($SD=0.08$).

There was, however, no study difference in the human students' learning; mean normalized gain, $M_{SE}=0.07$ ($SD=0.59$) vs. $M_{GS}=0.14$ ($SD=0.27$); $t(59)=-0.644$, $p=0.522$.

Depth of Learning

The strong correlation between tutee and tutor learning indicates that when the tutee commits shallow learning (which by definition shows good behavior at the surface level without actual learning gain), then the tutor might not learn well.

As mentioned earlier, one potential pit-fall for SimStudent's shallow learning in the APLUS environment is the structure of the Quiz. The likelihood of shallow learning would become higher when the Quiz problems are fixed. To test if the fixed set of Quiz problems actually induced SimStudent's shallow learning, and, if so, whether SimStudent's shallow learning also induced human students' shallow learning, we analyzed both human students' and SimStudent's shallow learning.

To test SimStudent's shallow learning, we investigated if SimStudents in the SE study who passed the (fixed) Quiz could also pass the (randomly generated) Quiz used in the GS study. There were 19 SimStudents who passed the Quiz in the SE study (SE passing SimStudent). We first extracted productions learned by those 19 SE passing SimStudents from the process data. Ten sets of the GS study Quiz (each with eight problems) were randomly generated. For each SE passing SimStudent, we then applied productions for each problem in the ten sets of the Quiz from the GS study.

Figure 2 shows the results of the SE passing SimStudents taking the GS study Quiz. The table shows the number of SE passing SimStudents that passed at most the specified number of quiz sets (out of 10).

There were 8 SE passing SimStudents who passed one or more sets of GS quiz. Only 7 (37%) SE passing SimStudents could pass two or more sets of GS Quiz. Interestingly, if SimStudent could pass two GS Quizzes, it could also pass all ten sets of GS Quizzes. To our surprise, 63% of SimStudents in the SE study passed the SE Quiz by committing "shallow" learning that was enough to pass the fixed set of Quiz problems.

Maximum Num. Quizzed Passed	0	1	10
Num. of SimStudent	11	1	7

Figure 2: Result of the SE SimStudents who passed the SE Quiz (N=19) taking the GS Quiz

Was SE SimStudent's shallow learning detrimental for human students learning? To see if human students actually committed shallow learning by quitting the tutoring session after seeing that their SimStudents passed the Quiz, we re-examined relationship between human students' learning and SimStudent's learning.

Figure 3 plots human student's plain test scores (Y-axis) and SimStudent's validation scores (X-axis). The figure only includes those 19 students who passed the Quiz in the SE Study. Data taken from a single human student are plotted as two dots connected with a vertical line. A large and a small dot show a human student's post- and post-test scores, respectively, both on the Y-axis.[2] The vertical line represents the pre- to post-test gain, with an upward line showing a positive gain and a downward line negative gain. SimStudent's learning (measures as the validation score) is shown as the position of the connected dots on the X-axis.

If there were students who committed shallow learning, then we should see the pair of dots connected with a relatively short upward line (i.e., a small positive gain) or a downward line (i.e., a negative gain) in the lower left corner. The human students with relatively short lines in the top area are likely to be ceiling students, and SimStudents in the right half are not likely to have committed shallow learning.

As can be seen in Figure 3, there are a group of human students who have relatively short or downward line at a relatively low pre-test score. They are the students who managed to have their SimStudents pass the Quiz, but the students themselves achieved very little learning gain.

Impact of Tutee Error for Tutor Learning

How do tutees' errors help tutor learning? To answer this question, we probed the process data to quantify several tutoring activities related to error detection and correction, and tested their correlations with tutor learning.

On average, SimStudent made 3.3 errors per problem ($SD=2.4$). The number of errors made by SimStudent per problem was not correlated with tutor learning; $r(84)=-.012$, $p=0.92$. The average probability for SimStudent making an error, which was computed as a ratio of incorrect suggestions to all suggestions per problem and averaged for all problems aggregated across all SimStudents, was not correlated with tutor learning; $r(85)=-.087$, $p=0.429$.

On average, human students correctly detected SimStudent's errors 2.3 times per problem ($SD=1.9$). There was no correlation between the number of times human students

[2] Test scores can be negative, because a point was subtracted for a wrong selection on multiple-choice items.

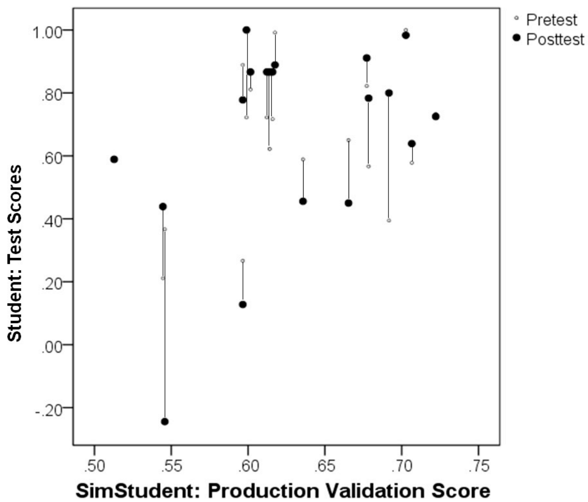

Figure 3: Relation between human students' test scores (Y-axis) and SimStudent's validation scores (X-axis). Large filled dots show post-test scores, whereas small circles show pre-test scores. Only students in the SE study who managed their SimStudents passing the Quiz are included.

correctly detected SimStudent's errors and tutor learning; $r(84)=-0.178$, $p=0.105$.

The tutors' probability of correctly detecting tutee errors (PED) significantly predicts tutors' normalized learning gain (NRG). The regression coefficient in predicting NRG with PED was NRG = 0.17 * PED, $p<0.05$. It is important to note that part of the test requires error detection within worked examples; success on the test therefore depends on this skill.

However, the NRG does not significantly predict the PED. It appears that tutor learning does not solely contribute to the ability of tutors to detect tutee errors. The regression coefficient in predicting PED with NRG was PED = 0.69 + 0.07*NRG, $p<.001$ and $p=0.17$, respectively.

As explained earlier, when a student provided negative feedback during tutoring, SimStudent asked the student to explain why the step was incorrect. The theory of self-explanation on erroneous examples predicts that responding to SimStudent's questions about its errors facilitates tutor learning (e.g., Grosse & Renkl, 2007). This is actually the case in our study as well. The ratio of "deep" explanations to all explanations on the "why am I wrong?" type of questions (DXP) was also significantly predictive of normalized gain (NRG). The regression coefficient was NRG = 0.44 * DXP with $p< 0.05$.

The above observations suggest that having tutors correctly detect tutee errors and elaborately explain the error would likely facilitate tutor learning. The current APLUS does not provide explicit assistance for the students to ensure such a good tutoring behavior. As some of the previous studies demonstrated (Leelawong & Biswas, 2008; Walker, et al., 2009), integrating a meta-tutor that guides the human student tutoring into APLUS might, therefore, improve the efficacy of APLUS.

Discussion

Shallow learning is an inevitable natural pathway for deep learning. When students engage in inductive learning, they usually search through a huge problem space with limited search heuristics that hardly avoid making errors. Indeed, there are very many different types of errors that students can make when doing induction (Matsuda, et al., 2009).

About 2/3 of SimStudents (12 out of 19) who passed the fixed Quiz in the SE Study were actually shallow learners (i.e., failed to pass the randomized Quiz in the GS study). On the other hand, there were no SimStudents who passed the Quiz in the GS study. Since there was a difference in the student's grade level for these two studies, the results must be interpreted with caution. Nonetheless, the GS Study data show a high correlation between SimStudent learning and human student learning. This means that the better SimStudent learns, the better the human student also learns. Therefore, a fixed set of quiz problems should work better than a randomized set of quiz problems, because it helps students tutor SimStudent, which makes a better SimStudent. In turn, this should lead to better tutor learning. In the SE Study, students who passed the Quiz used a higher percent of the failed Quiz problems for tutoring (M=0.95, SD=0.11) than students who did not pass (M=0.59, SD=0.42); $t(28)=-4.079$, $p=.000$, suggesting that copying the failed Quiz problems helped students in managing to pass the Quiz.

On the other hand, the high correlation between SimStudent's and (human) students' learning also suggests that failing to detect SimStudent's shallow learning is likely to cause students' poor learning. Therefore, the students must detect SimStudent's shallow learning. Our data show that catching SimStudent's shallow learning is rather inexpensive – only two sets of Quiz problems are enough. Having SimStudent take two or more different sets of Quiz problems should help students detect SimStudent's shallow learning.

Our current data also show that tutors learn from errors that the tutee makes. The probability of correctly detecting tutee errors is significantly predictive of tutor learning. Also, the ratio of elaborated explanations to all explanations given to incorrect steps is significantly predictive of tutor learning.

In conclusion, our data suggest that the inevitable nature of inductive learning, i.e., the tutee's *intermediate* shallow learning and errors of commission, facilitate tutor (as well as tutee) learning. In a certain situation (such as APLUS), letting the tutee reach shallow learning might help the tutor manage to teach the tutee without too much of a teaching burden. However, it is crucial for the tutor to detect the tutee's shallow learning and continue teaching toward deeper understanding. Our data also suggest that errors that the tutee makes during tutoring are beneficial both for the tutor and tutee. For the tutee, corrective feedback expedites its learning. For the tutor, elaborated reflective explanations on tutee errors facilitate learning.

The above findings also suggest that weaving fixed and randomly generated sets of quiz problems should induce optimal learning both for SimStudent and human students. One realization would be to provide a set of randomly generated Quiz problems and let SimStudent try the same (fixed) set of problems until SimStudent passes them, and then provide another set of randomly generated Quiz problems. As shown in Figure 2, passing only two sets of randomly generated Quizzes would be enough to ensure SimStudent's deep learning, which in turn prevents students from shallow learning.

Acknowledgments

The research reported here was supported by National Science Foundation Award No. DRL-0910176 and the Institute of Education Sciences, U.S. Department of Education, through Grant R305A090519 to Carnegie Mellon University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. This work is also supported in part by the Pittsburgh Science of Learning Center, which is funded by the National Science Foundation Award No. SBE-0836012.

References

Biswas, G., Jeong, H., Kinnebrew, J. S., Sulcer, B., & Roscoe, R. (2010). Measuring Self-Regulated Learning Skills through Social Interactions in a teachable Agent Environment. *Research and Practice in Technology Enhanced Learning*, 123-152.

Biswas, G., Leelawong, K., Schwartz, D., Vye, N., & Vanderbilt, T. T. A. G. a. (2005). Learning by teaching: a new agent paradigm for educational software. *Journal Applied Artificial Intelligence, 19*(3&4), 363-392.

Chase, C., Chin, D., Oppezzo, M., & Schwartz, D. (2009). Teachable Agents and the Protégé Effect: Increasing the Effort Towards Learning. *Journal of Science Education and Technology, 18*(4), 334-352.

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*(1), 1-35.

Grosse, C. S., & Renkl, A. (2006). Learning from worked examples: what happens if errors are included? . In J. E. P. Gerjets, R. Joiner & P. Kirschner (Eds.), *Instructional design for effective and enjoyable computer-supported learning*. Tubingen: Knowledge Media Research Center.

Grosse, C. S., & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction, 17*(6), 612-634.

Koedinger, K. R., Baker, R. S. J. d., Cunningham, K., Skogsholm, A., Leber, B., & Stamper, J. (2010). A Data Repository for the EDM community: The PSLC DataShop. In C. Romero, S. Ventura, M. Pechenizkiy & R. S. J. d. Baker (Eds.), *Handbook of Educational Data Mining*. Boca Raton, FL: CRC Press.

Leelawong, K., & Biswas, G. (2008). Designing Learning by Teaching Agents: The Betty's Brain System. *International Journal of Artificial Intelligence in Education, 18*(3).

Matsuda, N., Cohen, W. W., Sewall, J., Lacerda, G., & Koedinger, K. R. (2008). Why Tutored Problem Solving may be better than Example Study: Theoretical Implications from a Simulated-Student Study. In B. P. Woolf, E. Aimeur, R. Nkambou & S. Lajoie (Eds.), *Proceedings of the International Conference on Intelligent Tutoring Systems* (pp. 111-121). Heidelberg, Berlin: Springer.

Matsuda, N., Keiser, V., Raizada, R., Yarzebinski, E., Watson, S., Stylianides, G. J., et al. (2012). Studying the Effect of Tutor Learning using a Teachable Agent that asks the Student Tutor for Explanations. In M. Sugimoto, V. Aleven, Y. S. Chee & B. F. Manjon (Eds.), *Proceedings of the International Conference on Digital Game and Intelligent Toy Enhanced Learning (DIGITEL 2012)* (pp. 25-32). Los Alamitos, CA: IEEE Computer Society.

Matsuda, N., Lee, A., Cohen, W. W., & Koedinger, K. R. (2009). A Computational Model of How Learner Errors Arise from Weak Prior Knowledge. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the Annual Conference of the Cognitive Science Society* (pp. 1288-1293). Austin, TX: Cognitive Science Society.

Matsuda, N., Yarzebinski, E., Keiser, V., Raizada, R., Stylianides, G., Cohen, W. W., et al. (2011). Learning by Teaching SimStudent – An Initial Classroom Baseline Study comparing with Cognitive Tutor. In G. Biswas & S. Bull (Eds.), *Proceedings of the International Conference on Artificial Intelligence in Education* (pp. 213-221): Springer.

Matsuda, N., Yarzebinski, E., Keiser, V., Raizada, R., Stylianides, G., & Koedinger, K. R. (2012). Motivational factors for learning by teaching: The effect of a competitive game show in a virtual peer-learning environment. In S. Cerri & W. Clancey (Eds.), *Proceedings of International Conference on Intelligent Tutoring Systems* (pp. 101-111). Heidelberg, Berlin: Springer-Verlag.

Pareto, L., Arvemo, T., Dahl, Y., Haake, M., & Gulz, A. (2011). A Teachable-Agent Arithmetic Game's Effects on Mathematics Understanding, Attitude and Self-efficacy. In G. Biswas, S. Bull, J. Kay & A. Mitrovic (Eds.), *Proceedngs of the International conference on Artificial Intelligence in Education* (pp. 247-255). Heidelberg, Berlin: Springer.

Siegler, R. S. (2002). Microgenetic studies of self-explanation. In N. Granott & J. Parziale (Eds.), *Microdevelopment: Transition processes in development and learning* (pp. 31-58). New York, NY: Cambridge University Press.

Walker, E., Rummel, N., & Koedinger, K. R. (2009). Integrating collaboration and intelligent tutoring data in the evaluation of a reciprocal peer tutoring environment. *Research and Practice in Technology Enhanced Learning, 4*(3).