Journal of Educational Psychology
2013, Vol. 105, No. 4, 1152–1163

0022-0663/13/$12.00 DOI: 10.1037/a0031955

Cognitive Anatomy of Tutor Learning: Lessons Learned With SimStudent

Noboru Matsuda, Evelyn Yarzebinski,
Victoria Keiser, Rohan Raizada,
and William W. Cohen
Carnegie Mellon University

Gabriel J. Stylianides
University of Oxford

Kenneth R. Koedinger
Carnegie Mellon University

This article describes an advanced learning technology used to investigate hypotheses about learning by teaching. The proposed technology is an instance of a teachable agent, called *SimStudent*, that learns skills (e.g., for solving linear equations) from examples and from feedback on performance. SimStudent has been integrated into an online, gamelike environment in which students act as "tutors" and can interactively teach SimStudent by providing it with examples and feedback. We conducted 3 classroom "in vivo" studies to better understand how and when students learn (or fail to learn) by teaching. One of the strengths of interactive technologies is their ability to collect detailed process data on the nature and timing of student activities. The primary purpose of this article is to provide an in-depth analysis across 3 studies to understand the underlying cognitive and social factors that contribute to tutor learning by making connections between outcome and process data. The results show several key cognitive and social factors that are correlated with tutor learning. The accuracy of students' responses (i.e., feedback and hints), the quality of students' explanations during tutoring, and the appropriateness of tutoring strategy (i.e., problem selection) all positively affected SimStudent's learning, which further positively affected students' learning. The results suggest that implementing adaptive help for students on how to tutor and solve problems is a crucial component for successful learning by teaching.

Keywords: learning by teaching, machine learning, SimStudent, teachable agent, tutor learning

It has been widely observed that students learn by teaching others (e.g., E. G. Cohen, 1994). Such an effect of learning by teaching (also known as the *tutor-learning effect*) has been empirically confirmed in many different domains for many different structures of peer tutoring with different ages and achievement levels (Roscoe & Chi, 2007). Despite a long-standing history of empirical studies on the tutor-learning effect, not enough is known about the cognitive and social theory of when, how, and why tutors learn (or fail to learn) by teaching.

A primary challenge in theory development for tutor learning is a lack of the *process data*, that is, a detailed record of interactions between tutors and tutees. Collecting rich process data from peer tutoring sessions can enable descriptions of tutoring activities at a fine level of granularity, such as dialogue between the tutor and the tutee, response accuracy, and timing and sequencing of actions. When combined with outcome data (e.g., test scores), this detailed information can allow further exploration of elements of cognitive and social theory of tutor learning. However, such process data are rarely available. Roscoe and Chi (2007) reported that only six out of thousands of related articles report both outcome and process data. An obvious reason for the lack of process data is the difficulty in collecting such data during a study in which human students tutor their peers.

In their meta-analysis of prior research, Roscoe and Chi (2007) summarized potential flaws in program design and implementation that might have impacted the tutor-learning effect. One way to avoid such flaws is to better understand the process of tutor learning and to provide appropriate facilities for the tutors. Knowledge gained from combined process and outcome data can aid iterative design engineering of more effective learning by tutoring.

To help advance the cognitive and social theory of tutor learning, we have developed a synthetic pedagogical agent as a tutee that students can interactively tutor. Such a pedagogical agent is often called a *teachable agent*, which in our case is named *SimStudent* (Matsuda, Cohen, Sewall, Lacerda, & Koedinger, 2007). SimStudent engages in genuine machine learning to learn proce-

This article was published Online First September 9, 2013.

Noboru Matsuda, Evelyn Yarzebinski, Victoria Keiser, Rohan Raizada, and William W. Cohen, School of Computer Science, Carnegie Mellon University; Gabriel J. Stylianides, Department of Education, University of Oxford, Oxford, England; Kenneth R. Koedinger, School of Computer Science, Carnegie Mellon University.

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 Grant R305A090519 awarded 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 National Science Foundation Award No. SBE-0836012.

Correspondence concerning this article should be addressed to Noboru Matsuda, School of Computer Science, Human-Computer Interaction Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213. E-mail: Noboru.Matsuda@cs.cmu.edu

dural problem-solving skills. SimStudent has been integrated into an online, gamelike environment called *APLUS* (Artificial Peer Learning environment Using SimStudent).

With APLUS and SimStudent, we have conducted three tightly controlled in vivo studies for middle-school students learning algebra linear equations (Matsuda, Cohen, et al., 2012; Matsuda et al., 2011; Matsuda, Yarzebinski, et al., 2012). Solving linear equations is a critical area in the early algebra curriculum, yet many secondary school students experience great difficulty making the transition from arithmetic to algebra, especially in learning how to solve equations (see, e.g., Bednarz & Janvier, 1996; Filloy & Rojano, 1989; Kieran, 1992; Linchevski & Herscovics, 1996). Developing an effective intervention to learn equation solving thus has an urgent, practical need as well.

In this article, we investigated the following research questions:

Question 1: Does SimStudent actually learn how to solve equations when tutored by students in an authentic classroom setting? Accordingly, do students learn by teaching SimStudent?

Question 2: How do tutor and tutee learning correlate with each other?

Question 3: When and how do students learn or fail to learn by teaching SimStudent?

To answer these questions, we conducted in-depth analyses across three in vivo studies to understand the underlying cognitive and social factors that contribute to tutor learning. These analyses benefited from both outcome and process data.

In the rest of the article, we first provide a survey of prior research on the tutor-learning effect and the teachable agent technology. We then introduce SimStudent and APLUS, with a technical overview of how SimStudent acts as a teachable agent. Next, we explain how students interactively tutor SimStudent and provide an overview of the data analysis, which includes empirical data collected from the three in vivo studies. We then discuss how and when students learn or fail to learn by teaching SimStudent based on the process and outcome data. We conclude with a discussion of directions for future research based on the lessons learned from our studies.

The Tutor-Learning Effect

The tutor-learning effect has been studied for many years (Chi, Siler, Jeong, Yamauchi, & Hausmann, 2001; P. A. Cohen, Kulik, & Kulik, 1982; Devin-Sheehan, Feldman, & Allen, 1976; Gartner, Kohler, & Riessman, 1971; Graesser, Person, & Magliano, 1995) and for different age groups, varying from elementary (Sharpley, Irvine, & Sharpley, 1983) to middle school (Jacobson et al., 2001; King, Staffieri, & Adelgais, 1998) to college (Annis, 1983; Topping, 1996). It has also been observed in various subject domains, including mathematics, reading, science, and social studies (P. A. Cohen et al., 1982; Cook, Scruggs, Mastropieri, & Casto, 1986; Mastropieri, Spencer, Scruggs, & Talbott, 2000; Mathes & Fuchs, 1994; Rohrbeck, Ginsburg-Block, Fantuzzo, & Miller, 2003), and in different forms of tutoring, including reciprocal tutoring (Palincsar & Brown, 1984), collaborative passage learning (Bargh & Schul, 1980), and small-group learning as opposed to peer-to-peer learning (Webb & Mastergeorge, 2003). It has also been demonstrated that tutors can learn by just preparing for teaching (Biswas et al., 2001).

Learning by teaching has been shown to be effective for minority populations. Robinson, Schofield, and Steers-Wentzell (2005) found that African American student tutors learned more from math peer tutoring than White students. Rohrbeck et al. (2003) found a larger effect size in groups with more than 50% minority enrollment than groups with lower minority enrollment. Other researchers found positive outcomes for students from underprivileged backgrounds (Greenwood, Delquadri, & Hall, 1989; Jacobson et al., 2001) and students with learning disabilities (Cook et al., 1986; Mastropieri et al., 2000).

Despite the fact that many experimental studies support the tutor-learning effect, the actual effect size has been known to be rather moderate (P. A. Cohen et al., 1982; Cook et al., 1986; Mastropieri et al., 2000; Mathes & Fuchs, 1994; Rohrbeck et al., 2003). The tutor-learning effect has been shown to be relatively more effective in math than reading. For example, P. A. Cohen et al. (1982) showed an effect size of .62 for math and .21 for reading, and Cook et al. (1986) showed an effect size of .67 for math and .30 for reading.

In sum, learning by teaching has the potential to be a successful intervention for a wide variety of student populations across many disciplines. It also has the potential to minimize the achievement gap between student demographic diversities. Despite the popularity of the tutor-learning effect, we lack an adequate cognitive theory of tutor learning. Understanding the underlying cognitive principles of tutor learning could facilitate the development of effective learning technologies and may improve on the rather small effect size of tutor learning.

Teachable Agent

There are a number of advantages of using a teachable agent technology to study the tutor-learning effect (e.g., VanLehn, Ohlsson, & Nason, 1994). First, it enables implementation of tight, precisely determined control conditions. For example, the variance of tutees can be controlled by having students teach the same version of the teachable agent. The teachable agent technology also allows researchers to control the competency of the tutee to see how it may affect tutor learning. Second, the teachable agent allows researchers to conduct peer-tutoring studies without the risk of harming tutees. Although nonexpert tutors have a greater chance of teaching inaccurate knowledge, previous studies showed that tutors often learned at the cost of tutee errors. Walker, Rummel, and Koedinger (2009) found that the amount of tutee errors had a significant positive correlation with tutor learning, whereas it had a significant negative correlation with tutee learning. Third, the teachable agent technology facilitates the collection of detailed process data showing interactions between the student and the agent, which is a major contribution of the current article.

There have been three major techniques used to build teachable agents: (a) Some teachable agents (TAs) solve problems using the shared knowledge that students create. Students using such *knowledge-sharing TAs* are often told that they teach the agent by directly providing the shared knowledge to the agent. For example, students teach Betty's Brain by drawing a concept map representing causal relationships between factors related to river ecology (Biswas, Leelawong, Schwartz, Vye, & The Teachable Agents Group at Vanderbilt, 2005; Leelawong & Biswas, 2008). (b) Another type of TA applies to the knowledge-tracing technique that Cognitive Tutors use to diagnose students' competency (Ritter, Anderson, Koedinger, & Corbett, 2007). Such *knowledge-tracing TAs* are equipped with a set of skills to be learned. Some of the skills are set to be inactive at the beginning to provide the agent with limited competency to solve problems. As the student tutors the agent, the model-tracer identifies the skill that was tutored and activates the tutored skill so that the agent can apply it to future problems. Pareto, Arvemo, Dahl, Haake, and Gulz (2011) developed a knowledge-tracing TA for students to learn arithmetic concepts. (c) The last type of TA integrates machine-learning engines that allow the TA to learn skills dynamically, arguably more accurately reflecting the tutor–tutee interaction. As an example of such a *knowledge-learning TA*, Michie, Paterson, and Hayes (1989) developed the Math Concept Learning System with an inductive logic programming engine (called ID3) developed by Quinlan (1986) to induce rules from examples, which enabled it to learn math skills and solve equations. STEP (Simulated, Tutorable Physics Student) is another example of the knowledge-learning TA in Physics (Ur & VanLehn, 1995).

SimStudent is an example of a knowledge-learning TA, but has several distinctive characteristics compared with other TAs. First, SimStudent is one of a few TAs that have been intensively used in authentic classroom settings. Other such empirically well-validated agents include Betty's Brain (Biswas, Jeong, Kinnebrew, Sulcer, & Roscoe, 2010) and the TA developed by Pareto et al. (2011). Second, in contrast to other TAs, which have been largely implemented in declarative domains, SimStudent learns algebra content with a focus on procedural problem solving. Third, SimStudent is an instance of a TA with a *humanlike* learning capability (Li, Matsuda, Cohen, & Koedinger, 2011; Ohlsson, 2008). SimStudent performs inductive learning to interactively generalize examples provided by the student. Therefore, a naturalistic tutoring dialogue can occur between the student and SimStudent. Fourth, because SimStudent inductively learns skills from examples, it may learn skills incorrectly, depending on the prior knowledge it is given and the way the student tutors SimStudent. One such common source of incorrect learning stems from ambiguities in examples. To the best of our knowledge, SimStudent is the first TA that models students' incorrect *learning*. Because students generally learn both from correct and incorrect examples (Booth & Koedinger, 2008), observing a TA learning incorrectly may positively impact tutor learning.

Overview of the Data Analysis

To connect the outcome and process data to advance cognitive and social theories of the tutor-learning effect, data from three in vivo classroom studies have been analyzed to address the three research questions mentioned in the introduction. To measure SimStudent's learning, we used the process data showing how well SimStudent performed on the quiz. To measure students' learning (i.e., tutor learning), we used test scores as the outcome data. The correlation between SimStudents' and students' learning was analyzed using these two variables as well. We focused on a number of factors in the process data to analyze how and when SimStudents' and students' learning play out.

The Learning Environment: APLUS and SimStudent

Figure 1 shows an example screenshot of APLUS with SimStudent. The SimStudent avatar is visualized in the lower left corner. There have been three versions of SimStudent developed with different avatar images, as shown in Figure 2. Different versions of SimStudent have different functionalities to address different research questions as described later.

The initial version of SimStudent is called *Lucy* and is represented as a single static image (see Figure 2 i). The second version of SimStudent is called *Stacy* (see Figure 2 ii) and is capable of three facial expressions, including a thinking pose when SimStudent commits to learning, a happy expression when a problem is solved, and a neutral expression otherwise. The third version of SimStudent is called *Tomodachi* (see Figure 2 iii). Students can customize Tomodachi's avatar by changing the name, hairstyle, skin color, eyes, and shirt. Tomodachi is capable of the same three facial expressions as Stacy.

Overview of Tutoring Interaction

In APLUS, a student interactively tutors SimStudent with the following tutoring actions:

• Pose a problem for SimStudent to solve in the Tutoring Interface. In Figure 1a, the student entered "$3x - 7$" and "11" in the first row of the equation table. SimStudent then attempts to solve the problem by applying learned productions and asking the student about the correctness of each step.

• Provide flagged (yes/no) feedback to SimStudent that shows the student's judgment on the correctness of SimStudent's steps. When the student provides negative feedback, SimStudent may make another attempt. In Figure 1a, SimStudent entered "18" in the second row, and asked whether the student thought it was a good move. The student then provided positive feedback.

• Provide help on what to do next. When SimStudent does not know what to do, SimStudent asks the student for help. To respond to the help request, the student demonstrates the next step in the tutoring interface. In Figure 1a, SimStudent got stuck after entering "18." In response, the student tutors SimStudent by showing it a possible next step, in this case entering "divide 3" for the transformation of the second row.

• Quiz SimStudent to gauge learning. Students may have SimStudent take (and retake) the quiz at any time during tutoring (see Figure 1c). Further details of the quiz are below.

There are also resources for students to review learning objectives in the unit overview and to review problem-solving procedures by studying worked-out examples. Clicking the different [Example] tabs displays complete examples in the Tutoring Interface. The [Unit Overview] tab provides a brief overview of the target unit (i.e., equations with variables on both sides), a model solution with elaborated explanations, and suggested problems for students to use when tutoring SimStudent.

Quiz

In the classroom studies, students were told that their goal was to tutor SimStudent well enough so that SimStudent would pass a predefined quiz. The quiz has four sections each with two equation problems. There is a one-step equation (e.g., $3x = 6$), three

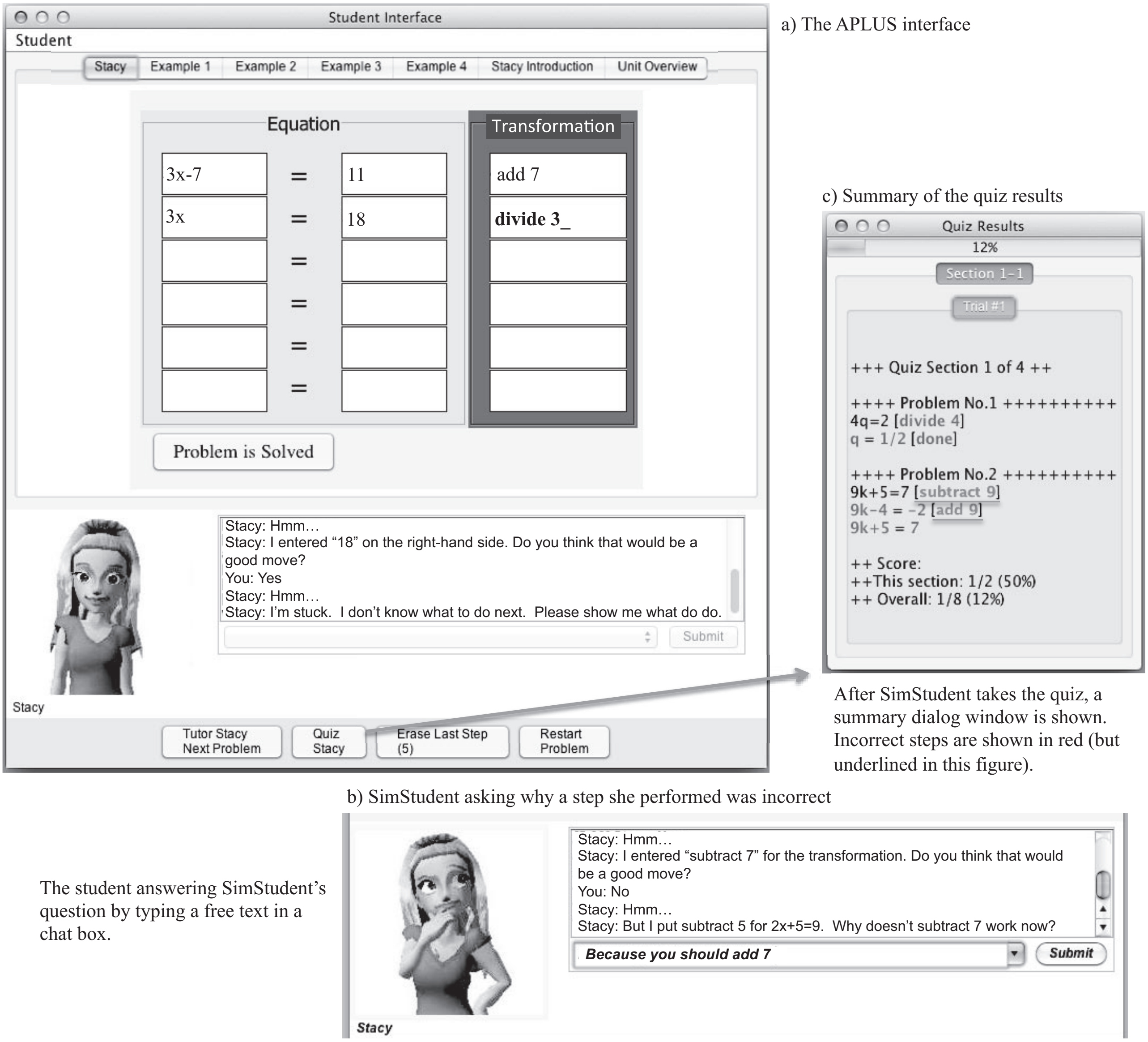

Figure 1. A screenshot of the Study II APLUS, the online gamelike learning environment in which the student can interactively tutor SimStudent. Students enter a problem on the first row of the Equation column. SimStudent attempts to solve the problem by entering steps (e.g., "$3x$" and "18" in this case). SimStudent asks the student about the correctness of the steps. When SimStudent cannot perform a step correctly, it asks the student for help. In this example, the student entered "divide 3" in the second row as a next step after SimStudent entered "18." SimStudent occasionally asks questions (b). In this example, SimStudent is asking for a reason why a step it performed is considered to be wrong. Students may have SimStudent take the quiz by clicking on the [Quiz Stacy] button. After SimStudent takes the quiz, a summary dialog window is shown (c). APLUS = Artificial Peer Learning environment Using SimStudent.

two-step equations (e.g., $-2x + 5 = 11$), and four equations with variables on both sides (e.g., $3 - 2x = 5x + 7$). SimStudent takes the quiz section by section, and must correctly solve both problems in each section to proceed to the next section.

After SimStudent takes the quiz, the overall results and correctness of the steps are displayed in a different window, as shown in Figure 1c. An embedded Cognitive Tutor Algebra program (Ritter et al., 2007) grades the quiz results. The Cognitive Tutor is invisible to students.

The quiz problems were randomly ordered for Study I, but they were ordered on the basis of increasing difficulty level for Studies II and III. The quiz problems were fixed for Studies I and II; that is, SimStudent was given the same set of quiz problems each time the student administered a quiz. For Study III, the quiz problems were generated on the fly while keeping the *type* of problems intact. This means that although the numbers and variables letters were changed each time SimStudent took the quiz, the positive and negative signs were pre-

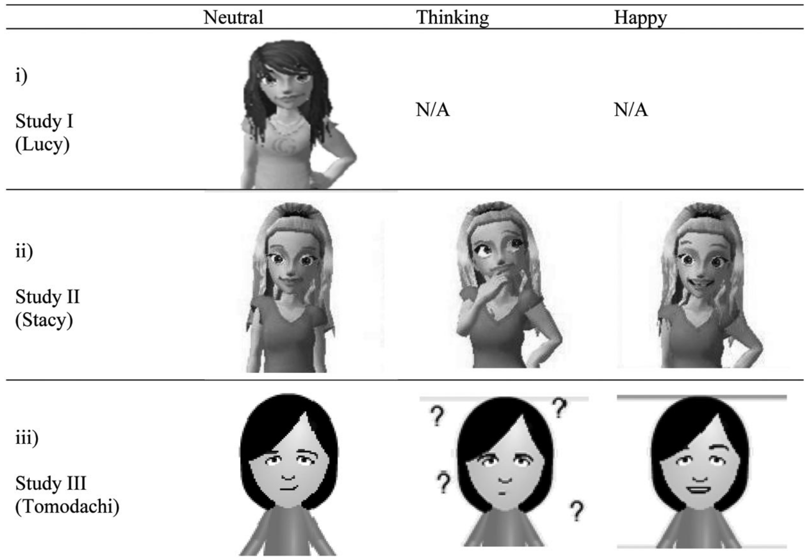

Figure 2. SimStudent's avatar image used in the three studies. There was only one static image used for Study I (Lucy). There are three facial expressions for Study II (Stacy): (a) neutral when waiting for the next problem to be entered, (b) thinking, and (c) happy after solving a problem. Study III (Tomodachi), which can be customized and uniquely named by each student, also has these three expressions. N/A = not applicable.

served. For example, $-3x + 5 = -7$ and $-2y + 1 = -10$ are considered to be isomorphic equations.

Self-Explanation

In Studies II and III, SimStudent had an ability to occasionally ask questions about things that students did, and these questions were intended to elicit students' self-explanations (Matsuda, Cohen, et al., 2012).

SimStudent's questions appear in the chat box at the bottom of the APLUS interface. SimStudent randomly selects a question from among a set of two to three questions that are relevant to each of three specific situations:

1. When the student inputs a *new problem* in the system, SimStudent asks why the student selected that problem or what that problem will help it learn.
2. If the student provides *negative feedback* on a step that SimStudent performed, SimStudent may or may not ask a question. If SimStudent has alternative actions to perform, it will not ask for an explanation. In cases in which SimStudent does solicit an explanation, it takes the last attempt that was made for the particular skill and asks why the step was incorrect, or how the situation is different from a previous step on which the same operation was used correctly.
3. SimStudent also asks questions after the student has provided a *hint* about transformation steps, not the results of the transformations (as the latter involves arithmetic calculation, and is thus often obvious). SimStudent will not ask a question at this point if it already asked about the student's negative feedback on the same step.

The ways students input their response varies depending on the type of question. For a question on a demonstrated hint or new problem, there is a drop-down menu available with prewritten, context-specific explanations. We hypothesized that such menu items would work as examples for students to learn (cf. Aleven & Koedinger, 2002). For questions about a demonstrated hint, the menu items use terminology such as *variable*, *constant*, and *coefficient* in a manner that reinforces their meanings. For questions about a new problem, the menu items include the key target concepts such as "It will help you learn how to deal with variables on both sides." Even when selecting an answer from the drop-down menu, students can also edit the selected text with their own words. For questions about negative feedback, for example, "Why is (x) wrong?" students need to input their own answers. Figure 1b shows an example of a student's response for SimStudent's question about why "subtract 7" is wrong for first transformation (which, by the way, is an example of SimStudent making an error that students commonly make).

SimStudent waits for student input before continuing to the next step of the equation. After the student clicks the *submit* button, the answer appears in the chat box below SimStudent's question. This explanation is also logged, but the answer does not affect SimStudent's learning. If the student clicks the *submit* button without providing an explanation, the student has essentially ignored SimStudent's question, and it will move on to the next step. In the classroom study, the students were not informed that they could skip the questions.

Overview of SimStudent's Learning

The underlying machine-learning paradigm used for SimStudent is a technique called *programming by demonstration* (Lau & Weld, 1998) that generalizes positive and negative examples to generate a set of hypotheses using a given set of background knowledge sufficient to interpret (or "explain") the examples. The positive and negative examples are provided by students as feedback and hints, as described in the previous section. Affirmative feedback (i.e., "yes") and hints become positive examples, whereas instances of negative feedback (i.e., "no") become negative examples.

SimStudent generalizes from these positive and negative examples and generates a set of production rules that can reproduce all positive examples but no negative examples. Each production represents *where* to focus attention to know *when* and *how* to apply a particular skill. SimStudent uses hybrid AI techniques to learn the where, when, and how parts of a production rule. Providing technical details of the learning algorithm is beyond the scope of this article, but can be found elsewhere (Matsuda et al., 2007).

As mentioned earlier, one of the unique characteristics of SimStudent is its ability to learn skills incorrectly. We hypothesize that students learn incorrect skills by making inappropriate inductions from examples due to inappropriate background knowledge (Matsuda, Lee, Cohen, & Koedinger, 2009). Such incomplete background knowledge allows students to rely on shallow problem-solving features instead of deep domain principles.

As an example, suppose that a student is about to generalize an example of "subtracting 3 from both sides of $2x + 3 = 5$." The student may recognize "+" in the left-hand side as the arithmetic operator instead of the sign of a term. As a consequence, the student may generalize this example to "subtract a number that follows an operator." Students who perceive such a shallow feature would also be likely to subtract 4 from both sides of $3x - 4 = 6$ as well, which is one of the most frequently observed student errors (Booth & Koedinger, 2008).

To model this type of incorrect learning, we "weakened" SimStudent's background knowledge by dropping the concept of an algebraic term in an expression and adding more perceptually grounded background knowledge, such as "get a number after an arithmetic operator." In a prior study (Matsuda et al., 2009), we validated the cognitive fidelity of SimStudent's learning by comparing SimStudent's and human students' learning. The study showed that SimStudent with "weak" prior knowledge learned skills incorrectly in a humanlike manner and generated humanlike errors when solving problems using the learned productions.

SimStudent applies learned productions to solve problems posed by a student, but the productions are not visible to the student. Therefore, the cognitive fidelity mentioned above could better facilitate tutor learning, because the student must identify, understand, and remediate SimStudent's errors, which evoke or foster metacognitive tutoring skills and a deep understanding of the domain knowledge.

Method

Classroom Studies and Data Collection

The three in vivo studies were conducted as controlled randomized trials under the direct supervision of the Pittsburgh Science of Learning Center (LearnLab.org). Each study was conducted as a part of regular algebra classes. The studies used the same general format that involved 5 (Study I) or 6 (Studies II and III) days in the classroom. On the first day, all students took a pretest using an online test form (as described in the Measures section). After taking the pretest, students were randomly split into two groups and studied algebra equations using the assigned material for two (Study I) or three (Studies II and III) class periods (one class period per day). All students then took an online posttest on the following day. Finally, all students took an online delayed test 2 weeks after the posttest.

Study I: Initial classroom trial. The primary goal of Study I was to evaluate the effectiveness of SimStudent (Matsuda et al., 2011). The version of APLUS and SimStudent used in Study I behaved exactly as described in the previous section and is called *Baseline* hereafter. Algebra I Cognitive Tutor (Ritter et al., 2007) was used for the control condition.

Study II: Self-explanation effect. In Study II, we focused on the self-explanation hypothesis, which conjectures that the tutor-learning effect is facilitated when the students are asked to explain and justify their tutoring decisions (Matsuda, Cohen, et al., 2012). To test this hypothesis, we compared SimStudent that did (the self-explanation condition) and did not (the baseline condition) ask questions.

Study III: Game show effect. For Study III, we compared the effect of learning by teaching SimStudent in APLUS with and without a Game Show feature (Matsuda, Yarzebinski, et al., 2012). In the Game Show, a pair of SimStudents, each tutored by a different student, compete by solving problems posed by the students who tutored them. This study was conducted to test the motivation hypothesis that conjectures that the more students are engaged in tutoring, the more tutor learning would be facilitated. The students in the Game Show condition were told to obtain the highest score in the Game Show, instead of having Tomodachi pass the quiz, which was the goal for the students in the non-Game Show condition.

Because the scope of this article does not include the motivation hypothesis, we do not discuss details of Study III here. However, we include Study III in the following analysis, because the control condition of Study III used the same version of SimStudent that Study II used for the self-explanation condition. Namely, the Study III SimStudent occasionally prompted students for explanations and justifications.

Participants

There were two schools involved in Study I. One school had 30 Algebra I (Grade 8) and 34 Algebra II (Grade 9) students, and the other school had 40 Algebra I (Grade 8) students. Study II involved one school with 160 Algebra I students in Grades 8, 9, and 10. Study III was conducted at the same school as Study II, and 141 Algebra I students in Grades 7 and 8 participated in Study III. To avoid a confounding factor of familiarity with the study, we excluded the ninth- and 10-grade students who were likely to have been included in Study II.

There were a significant number of absentees in each study. For the analysis in the following sections, we included only students who took all three (pre, post, and delayed) tests and participated in all classroom sessions. As a consequence, the following analyses contain 33 (32%), 81 (51%), and 69 (49%) of students for Study I, II, and III, respectively.

Measures

Outcome of tutee learning. To quantify tutee learning (i.e., SimStudent's achievement), we use the number of quiz sections that SimStudent passed, which differed in format among the three studies.

Outcome of tutor learning. Students' learning was measured with online tests that consisted of two parts—the *Procedural Skill Test* (PST) and the *Conceptual Knowledge Test* (CKT). The tests had three isomorphic versions that were counterbalanced for pre-, post-, and delayed tests. Two test items were considered isomorphic when they were of identical type, but included different letters and numbers. Equations were carefully varied so that two isomorphic equations shared the same properties in their solutions (e.g., whole number vs. fraction).

The PST had three types of test items: (a) six equation-solving items. Students were asked to show their work on a piece of paper; (b) twelve agree/disagree items to determine whether a given operation was a logical next step for a given equation; (c) five worked-out items to identify the incorrect step in a given incorrect solution (multiple choice) and explain why (free response). The CKT had two types of test items: (d) thirty-eight true/false items asking about basic algebra vocabulary to identify constants, variables, and like terms; (e) ten true/false items to determine whether two given expressions are equivalent.

For Studies II and III, the following changes were made on the online test: (a) Four additional one-step equations were added to the equation-solving items. (b) A "Not Sure" option was added for multiple-choice items to lower the chance of students making random guesses. Students were told that they would lose a point for an incorrect answer for multiple-choice questions, but there was no penalty for selecting "Not Sure."

The test items were graded as follows. For the equation-solving items, students received a score of 1 if their answer was correct and partial credit based on their written work if their answer was incorrect. For the multiple-choice items, students received a score of 1 for a correct answer, 0 for "Not Sure," and -1 for an incorrect answer.

Cognitive and Social Factors of Interest

APLUS automatically collects detailed data showing the interaction between students and SimStudent with additional narratives such as the response correctness. In the current analysis, we focus on the following variables:

1. The accuracy of students' *feedback* and *hints*. The accuracy of *response* is an aggregation of feedback and hints.
2. The likelihood of responding to SimStudent's hint request, which is the ratio of hints provided by a student to the total number of hints requested by SimStudent. Although students must answer SimStudent's hint request to proceed to the next step, they sometimes avoided answering by starting a new problem or giving a quiz.
3. The frequency of self-explanations submitted by students during tutoring.
4. The type of problems tutored. Although students were explicitly told that SimStudent must be able to solve equations with variables on both sides to pass the quiz, they needed to start with easier types to work up to the target difficulty.
5. The degree of repetition in selecting problems for tutoring. Students in our studies often used quiz problems during tutoring. As mentioned before, the problems in the quiz were fixed for Study II, but only the *type* of problems was fixed for Study III. To avoid confusion, we shall use the term *problem* to mean the *exact* same problem for Study II and the same *type* of problem for Study III. The *problem repetition ratio* is then the ratio of the number of problems tutored more than once to the total number of problems tutored.
6. Time on task. The amount of time students spent tutoring problems and giving explanations to SimStudent. This time does not include the quiz or the resource usage.
7. Tutor's prior knowledge, that is, each student's PST and CKT pretest scores.
8. Tutee's learning outcome, that is, the number of quiz sections that SimStudent passed.

Results

This section is organized to answer the three major research questions mentioned in the introduction. We first show results about SimStudent's and students' learning outcomes addressing the first research question. We then show the correlations between tutor and tutee learning that answers the second research question. Finally, we show major findings obtained from the process data showing the cognitive and social factors that have significant influence on tutor and tutee learning.

Learning Outcomes

Because the three studies were conducted at different schools in different years, we first tested whether there was any population difference among the three studies. A one-way analysis of variance (ANOVA) was conducted with the independent variable of study (I, II, III) and the dependent variable of pretest score aggregated across two conditions. For both PST and CKT, the mean pretest score for Study I was significantly higher than Study II, which was significantly higher than Study III; for PST, $F(2, 180) = 23.58$, $p < .001$; for CKT, $F(2, 180) = 44.81$, $p < .001$. The difference in the pretest scores might reflect the age difference between the studies. Study III had the youngest student population.

Tutee-learning outcome: Performance on the quiz. Figure 3 shows the number of students whose SimStudent passed the quiz during the intervention. In Study I, none of the 18 students in the SimStudent condition managed to get their SimStudent to pass all four sections of the quiz. Only five students managed to pass quiz Section 1, and of those five, only one student passed quiz Section 3. For Study II, 36 out of 81 students managed to pass all four sections of the quiz within the allotted 3 days. Nearly all students (78 out of 81, i.e., 96%) passed at least Section 1. To our surprise, in the Study III baseline condition, we again observed that none of the students managed to have their SimStudent pass the quiz. Only 22 out of 40 (55%) students passed quiz Section 1. As mentioned earlier, Study III involved younger students and showed lower pretest score than the other two studies. The students in Study III might have less prepared for tutoring.

Tutor-learning outcome: Test scores. A summary of the test scores is shown in Table 1. The table shows mean scores for the pre-, post-, and delayed tests for all three studies. No condition difference on the pretest was found both for the PST and the CKT across the three studies. We thus conducted a 2×3 repeated measures ANOVA, with condition (study vs. control) as a between-subjects variable and test time (pre, post, and delayed) as

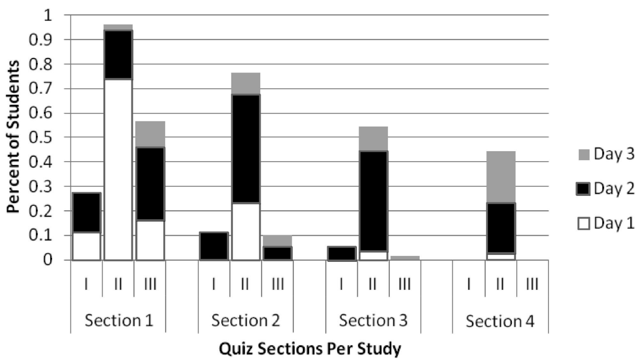

Figure 3. Number of students whose SimStudent passed the quiz. I, II, and III represents each study. Study I did not have a Day 3. The quiz had four sections with two equation problems per section, which were randomly ordered for Study I and ordered on the basis of difficulty level for Studies II and III. The quiz problems were fixed for Studies I and II. For Study III, the quiz problems were generated on the fly while keeping the type of problems intact.

a within-subjects repeated variable. The ANOVA was run on the PST and the CKT separately for each study.

For the PST, there was neither a main effect of condition nor an interaction effect between test time and condition. There was, however, *an iterative enhancement of the effect of tutor learning indicated as the gain on the test scores*. In Study I, the impact of test time was absent. In Study II, however, there was a main effect of test time, $F(2, 78) = 34.85$, $p < .001$. A further analysis showed that students' average test scores were significantly higher in the delayed test ($M = .68$, $SD = .27$) than both the pretest ($M = .54$, $SD = .26$; $p < .001$, $d = 0.53$) and the posttest ($M = .57$, $SD = .31$; $p < .001$, $d = 0.38$). The difference between the pre- and posttests was statistically indistinguishable. The reason for the higher delayed test scores in Study II might not be completely due to the study intervention. There were algebra classes between the post- and the delayed test (2 weeks apart) in which regular teachers continued teaching equation solving. For Study III, again, there was a main effect of test time, $F(2, 66) = 8.81$, $p < .001$. Both posttest ($M = .45$, $SD = .20$; $p < .01$, $d = .35$) and delayed test ($M = .46$, $SD = .23$; $p < .001$, $d = .37$) were significantly higher than pretest ($M = .38$, $SD = .20$). The difference between the post- and delayed test was not statistically significant.

For the CKT, there was neither a main effect of test time nor a main effect of condition for all three studies.

Correlation Between Tutee and Tutor Learning

In Study III, tutee learning (the number of quiz sections that SimStudent passed) has a significant correlation with tutor learning (the normalized gain on the PST); $r(39) = .37$, $p < .05$. There was no significant correlation between tutee and tutor learning for Studies I and II.

Cognitive and Social Factors for Tutee and Tutor Learning

What affected tutee learning? It is surprising to observe that so many students failed to sufficiently tutor SimStudent to pass the quiz. To understand why, we conducted comparative analyses by splitting students into two groups based on the median quiz progress. For this analysis, we included students from both conditions in Study II ($N = 81$) and those in the control condition in Study III, in which the goal of tutoring was to have SimStudent pass the quiz ($N = 40$). Study I students were excluded from this analysis, because the order of quiz items in Study I was not compatible with Studies II and III.

Students were split into the *successful group* and the *unsuccessful group* using the median of the quiz section passed. For Study II, the split occurred at Section 3 (successful $n = 44$ vs. unsuccessful $n = 37$), whereas for Study III, the split occurred at Section 1 (successful $n = 21$ vs. unsuccessful $n = 18$). Within each study, we compared the two groups for a number of factors using independent samples t tests. Table 2 shows the results of this analysis.

Table 1
Test Scores Summary

Study	Pre-test		Post-test		Delayed-test	
I						
	CogTutor	Baseline	CogTutor	Baseline	CogTutor	Baseline
P	.67 (.24)	.74 (.19)	.73 (.20)	.76 (.21)	.65 (.25)	.72 (.17)
C	.55 (.13)	.62 (.15)	.57 (.10)	.63 (.16)	.52 (.19)	.58 (.13)
II						
	Baseline	SelfExpl	Baseline	SelfExpl	Baseline	SelfExpl
P	.54 (.27)	.52 (.26)	.57 (.29)	.57 (.33)	.68 (.25)	.68 (.30)
C	.29 (.23)	.29 (.27)	.32 (.24)	.33 (.27)	.30 (.25)	.35 (.26)
III						
	SelfExpl	Game Show	SelfExpl	Game Show	SelfExpl	Game Show
P	.34 (.19)	.44 (.19)	.41 (.21)	.49 (.19)	.43 (.23)	.50 (.23)
C	.13 (.20)	.19 (.20)	.16 (.19)	.21 (.20)	.15 (.19)	.22 (.20)

Note. CogTutor = cognitive tutor; Baseline = the baseline Artifical Peer Learning environment Using SimStudent (APLUS) and SimStudent; P = Procedural Skill Test; C = Conceptual Knowledge Test; SelfExpl = APLUS and SimStudent with self-explanation prompt; Game Show = APLUS and SimStudent with the Game Show feature. Each cell shows the mean, with the standard deviation in parentheses.

Table 2
Comparison Between Successful and Unsuccessful Groups Based on a Median Split for Number of Quiz Sections Passed

Factor	Study	Quiz performance: Successful	Quiz performance: Unsuccessful	t	Cohen's d	df
Procedural Normalized Gain	II	.15 (.52)	.03 (.51)	−1.01	0.23	77[†]
	III	.22 (.21)	−.01 (.29)	−2.80**	0.92	37[†]
Correct Feedback	II	.86 (.07)	.83 (.08)	−1.52	0.34	79
	III	.78 (.06)	.68 (.14)	−2.80**	1.21	21[a]
Correct Hint	II	.76 (.15)	.63 (.15)	−3.73***	0.83	79
	III	.56 (.13)	.30 (.25)	−3.95***	1.61	24[b]
Disregarding Hint Requests	II	.44 (.16)	.52 (.20)	1.86	0.41	79
	III	.16 (.09)	.29 (.16)	3.14**	1.03	37[†]
Repeating Problem Element	II	.13 (.13)	.23 (.19)	2.86**	0.72	62[c]
	III	.38 (.12)	.46 (.19)	1.77	0.57	38

Note. Standard deviations appear in parentheses.
[a] Levene's test indicated unequal variances, ($F = 10.07$, $p < .01$); df adjusted from 38 to 21. [b] Levene's test indicated unequal variances, ($F = 12.68$, $p < .001$); df adjusted from 38 to 24. [c] Levene's test indicated unequal variances, ($F = 9.48$, $p < .01$); df adjusted from 79 to 62.
** $p < .01$. *** $p < .001$. † Maximum df is 79 for Study II and 38 for Study III. Numbers marked with this symbol had $79-N$ or $38-N$ cases removed on the basis of a z-score outlier analysis of ± 3.

First, the correctness of the student's *feedback* and *hints* had a notable influence on SimStudent's learning. For Study II, students in the successful group provided correct hints more often than students in the unsuccessful group. There was, however, no group difference in the accuracy of the feedback provided. For Study III, students in the successful group provided both correct hints *and* accurate feedback more often than students in the unsuccessful group.

Second, the likelihood of responding to SimStudent's hint request also has a notable difference. For Study III, the successful students responded to hint requests more often than unsuccessful students. The likelihood is, however, not significantly different between successful and unsuccessful groups for Study II.

Third, the problem repetition ratio was different. For Study II, the successful group tended to repeat the exact same problem less often than the unsuccessful group. For Study III, however, the difference was only marginal.

What affected tutor learning? In this analysis, we use the normalized gain of the PST from the pre- to the posttest as the measurement for the tutor learning. This analysis includes the same student data used in the tutee-learning analysis mentioned in the previous section.

First, the more the target problems were tutored (i.e., equations with variables on both sides), the more the *students* learned. This correlation was observed in both Study II, $r(79) = .26$, $p < .05$, and Study III, $r(39) = .36$, $p < .05$.

Second, the more the students gave self-explanations on the target problems, the more the students learned. Again, this correlation was observed in both Study II, $r(38) = .32$, $p < .05$, and Study III, $r(37) = .33$, $p < .05$.

Third, the more the students provided a correct tutoring responses (a combination of feedback and hint), the more the students learned, although this correlation was observed only in Study III, $r(38) = .31$, $p < .05$.

To our surprise, there was no correlation between the time on task and tutor learning in both studies: Study II, $r(81) = .09$, $p = .40$; Study III, $r(40) = .08$, $p = .61$.

Impact of Prior Knowledge for Tutor and Tutee Learning

We first show the impact of the tutee's prior knowledge on *tutor* learning. SimStudents for Study II (Stacy) and Study III (Tomodachi) were equally pretrained on more one-step equations than the SimStudent in Study I (Lucy). An independent samples t test confirmed that both Stacy and Tomodachi performed better on the first three tutoring problems than Lucy.

To see how the tutee's performance affected the tutor's performance, students' response accuracy was computed as a ratio of correct responses (i.e., feedback or hint) to all responses for each step in the first three tutored problems. On average, students in Studies II and III gave accurate responses more often than Study I students. Students in Studies II and III showed an average response accuracy of .76 ($SD = .21$), whereas students in Study I showed an average response accuracy of .57 ($SD = .26$). The difference is statistically significant, $t(134) = -3.49$, $p < .001$, $d = 0.60$.

One possible explanation for students' higher response accuracy in Studies II and III is that it is easier to recognize correct steps as correct than to identify incorrect steps as incorrect. Because Stacy and Tomodachi performed more steps correctly, the students in Studies II and III were able to correctly provide positive feedback more easily. When aggregated across all three studies, SimStudent's performance accuracy and students' response accuracy were actually highly correlated, $r(135) = .69$, $p < .001$.

Next, we analyzed the impact of the tutor's prior knowledge on tutor learning. The PST and CKT pretest were both predictive of students' posttest scores on the PST. A regression analysis with PST and CKT pretest scores as independent variables and the PST posttest score as a dependent variable revealed the following regression coefficients: PST_Post = .70 × PST_Pre + .12 × CKT_Pre + .17. An identical analysis was also conducted for the CKT posttest; a regression analysis with the PST and CKT pretest scores as independent variables and the CKT posttest score as a dependent variable revealed the following regression coefficients: CKT_Post = .25 × PST_Pre + .50 × CKT_Pre + .05.

Discussion

Results from the experiment provide four sets of important information to understand tutor learning. First, our data show that learning by teaching SimStudent is effective for learning procedural skills measured by the PST, as shown in Study III, but not for learning conceptual knowledge measured by the CKT.

Second, there is a significant correlation between tutee and tutor learning. Students tended to learn more when they tutored SimStudent correctly (i.e., with an accurate response) and appropriately (i.e., on appropriate problems with a sufficient amount of explanations).

Third, there were some notable differences in the way that the successful and the unsuccessful groups tutored SimStudent. Students in the unsuccessful group had trouble teaching SimStudent well, perhaps without even recognizing that they were not teaching appropriately. This manifested itself in students making many of the same errors, not properly responding to SimStudent's hint requests, and repeatedly teaching the same problem.

Fourth, both tutee and tutor's prior knowledge affected tutor learning. When the tutee had higher prior knowledge, the tutor tended to respond more accurately, which was further correlated with tutor learning. Our data also showed, however, that the tutor's prior competence both on conceptual and procedural knowledge was strongly predictive of tutor learning.

Finally, both SimStudent and APLUS have been iteratively improved from Study I to Study III, which may explain the gradual enhancement of the outcome. There was a population difference in the pretest score. Both for the PST and the CKT, students in Study I scored higher than the students in Study II, who outperformed the students in Study III. Yet, only Study III showed a significant gain in PST scores from pre- to posttest.

Tutor Help

Our findings show that learning by teaching does not happen automatically. Students need help to tutor SimStudent correctly and appropriately. Other research has also pointed out that students often do not correctly recognize their own misunderstandings (King, 1998). In our studies, students often unknowingly made inappropriate tutoring decisions and provided incorrect feedback and hints, which affected SimStudent's learning. These behaviors were negatively correlated with tutor learning.

One idea to provide such *tutor help* is to integrate a third agent (a *meta-tutor*) into the APLUS environment, a commonly used idea in the context of multiagent learning systems (e.g., Biswas et al., 2005; Vassileva, McCalla, & Greer, 2003). The meta-tutor oversees students' tutoring activities and provides them with just-in-time scaffolding.

The meta-tutor could provide students with both *cognitive* help regarding domain knowledge about how to solve problems and *metacognitive* help regarding proper tutoring methods. Some studies show that tutors can be trained to be a better tutor, which facilitates tutor learning (Ismail & Alexander, 2005; King et al., 1998). Other studies show the effect of the tutor help (Biswas et al., 2010; Walker et al., 2009), but none of them have explored the differing effects of cognitive and metacognitive help. It is therefore important to study how to implement cognitive and metacognitive help, how they foster tutor learning, and how well students learn tutoring skills from these different types of interactive support.

Another possibility is for the teachable agent itself to provide tutor help. For example, if a student poses the same problem (or the same type of problem) multiple times, then SimStudent could alert the tutor. The difference in the source of tutor help would have different social and affective impacts on the student. This might become particularly subtle when the student has established a different rapport with SimStudent and the meta-tutor. Studying the social factors of tutor help would therefore be important (Ogan, Finkelstein, Mayfield, D'Adamo, Matsuda, & Cassell, 2012).

The Effect of Self-Explanation for Tutor Learning

The current data show that the tutor-learning effect in APLUS is limited to procedural skills. Further studies will be needed to investigate the tutor-learning effect on conceptual knowledge. We hypothesized that self-explanations would facilitate learning conceptual knowledge, because good explanations contain conceptual justifications for algebraic operations. However, the students sometimes provided shallow responses (e.g., "Because you didn't add right") or irrelevant responses (e.g., "Because I just did"). The current version of SimStudent does not parse students' responses; instead, it simply proceeds to the next step. Empirical studies show that the tutee's questions have substantial influence on tutor learning (Roscoe & Chi, 2004). Thus, if SimStudent requested elaboration or further reflection on a given response, it may facilitate tutor learning. This kind of question is called a *reflective knowledge-building* question, and its effect has been well researched (Roscoe & Chi, 2007). Building such an intelligent teachable agent is therefore an important direction for future research (Carlson, Keiser, Matsuda, Koedinger, & Rose, 2012).

Learning by Teaching Versus Cognitive Tutoring

Our data show similarities and differences between learning by teaching and learning by cognitive tutoring (i.e., more direct instruction). The effect of self-explanation, for example, was evident for both styles of learning. Possession of prerequisite knowledge also has a notable influence on both styles of learning (Booth & Koedinger, 2008).

A notable difference between the two learning styles is the degree to which students can practice metacognitive skills. Our data show that the accuracy of tutoring responses, the frequency of self-explanations, and the type of problems tutored all positively correlate with tutor learning. To achieve successful learning, students must simultaneously monitor both their tutee's performances and their own. This double-edged monitoring requires more complicated metacognitive skills than solving problems alone in the context of cognitive tutoring.

There is also a difference in the timing of feedback. The feedback in the context of APLUS, that is, the system's reaction to the correctness of the student's tutoring activities, is delayed. The current version of APLUS does not provide students with any explicit feedback on their tutoring activities. Students later notice when they have made mistakes by reviewing the quiz summary or by observing SimStudent's undesired behaviors during tutoring. As an example of the second mistake, even when a student incorrectly demonstrated "subtract 4" for "$3x - 4 = 10$" with a correct intention to isolate the "$3x$" on the left-hand side, SimStudent might correctly suggest entering "$3x - 8$" for the left-hand side of the new equation, instead of "3*x*," which is what the student expected to see.

The gap between the student's expectations and SimStudent's actual performance might motivate students to reflect on their tutoring actions. This is a kind of "intelligent novice" model of desired performance (Mathan & Koedinger, 2005) for tutor learning. Embedding the above-mentioned tutor help into the model of desired tutor performance might thus facilitate tutor learning. Observing the tutee's performance is a distinctive form of learning from correct and incorrect examples available in learning by teaching.

Conclusion

Students learn by teaching others. Our data show that students learn by teaching primarily when they teach the target skills correctly and appropriately. The accuracy of students' responses (i.e., feedback and hints), the quality of students' explanations during tutoring, and the appropriateness of tutoring strategy (i.e., problem selection) all affected SimStudent's learning outcome, which further affected students' learning.

Students' prior knowledge has a strong influence on tutor learning. If students are not well prepared to tutor, the benefits of tutor learning might be reduced. Alternatively, once students become domain experts and can solve problems fluently (hence become better teachers), the benefit of tutor learning might also decline. Tutor learning is essentially a paradoxical phenomenon whose mechanisms have yet to be fully elucidated.

Students make errors when teaching and get stuck when providing hints, both of which are detrimental for tutor learning. Providing more tutor help in the form of cognitive and/or metacognitive support may be critical to optimizing tutor learning.

The competence of the tutee also affects tutor learning as well. Carefully designing SimStudent's learning ability and adaptively assigning an optimized SimStudent on the basis of the student's competency would further provide us with insight into successful learning by teaching.

References

Aleven, V., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science, 26,* 147–179. doi:10.1207/s15516709cog2602_1

Annis, L. F. (1983). The processes and effects of peer tutoring. *Human Learning, 2,* 39–47.

Bargh, J. A., & Schul, Y. (1980). On the cognitive benefits of teaching. *Journal of Educational Psychology, 72,* 593–604. doi:10.1037/0022-0663.72.5.593

Bednarz, N., & Janvier, B. (1996). Emergence and development of algebra as a poblem solving tool: Continuities and discontinuities with arithmetic. In N. Bednarz, C. Kieran, & L. Lee (Eds.), *Approaches to algebra* (pp. 115–136). Dordrecht, the Netherlands: Kluwer.

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, 5,* 123–152.

Biswas, G., Leelawong, K., Belynne, K., Viswanath, K., Vye, N. J., Schwartz, D. L., & Davis, J. (2004). Incorporating self-regulated learning techniques into learning by teaching environments. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (pp. 120–125). Mahwah, NJ: Erlbaum.

Biswas, G., Leelawong, K., Schwartz, D., Vye, N., & The Teachable Agents Group at Vanderbilt. (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence: An International Journal, 19,* 363–392. doi:10.1080/08839510590910200

Booth, J. L., & Koedinger, K. R. (2008). Key misconceptions in algebraic problem solving. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 571–576). Austin, TX: Cognitive Science Society.

Carlson, R., Keiser, V., Matsuda, N., Koedinger, K. R., & Rose, C. (2012). Building a conversational SimStudent. In S. Cerri & W. Clancey (Eds.), *Proceedings of the International Conference on Intelligent Tutoring Systems* (pp. 563–569). Heidelberg, Berlin: Springer-Verlag.

Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science, 25,* 471–533. doi:10.1207/s15516709cog2504_1

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64,* 1–35. doi:10.2307/1170744

Cohen, P. A., Kulik, J. A., & Kulik, C-L. C. (1982). Education outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal, 19,* 237–248.

Cook, S. B., Scruggs, T., Mastropieri, M., & Casto, G. (1986). Handicapped students as tutors. *Journal of Special Education, 19,* 483–492. doi:10.1177/002246698501900410

Devin-Sheehan, L., Feldman, R. S., & Allen, V. L. (1976). Research on children tutoring children: A critical review. *Review of Educational Research, 46,* 355–385. doi:10.2307/1170008

Filloy, E., & Rojano, T. (1989). Solving equations: The transition from arithmetic to algebra. *For the Learning of Mathematics, 9,* 19–25.

Gartner, A., Kohler, M., & Riessman, F. (1971). *Children teach children: Learning by teaching*. New York, NY: Harper & Row.

Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology, 9,* 495–522. doi:10.1002/acp.2350090604

Greenwood, C. R., Delquadri, J. C., & Hall, R. V. (1989). Longitudinal effects of classwide peer tutoring. *Journal of Educational Psychology, 81,* 371–383. doi:10.1037/0022-0663.81.3.371

Ismail, H. N., & Alexander, J. M. (2005). Learning within scripted and nonscripted peer-tutoring sessions: The Malaysian context. *The Journal of Educational Research, 99,* 67–77. doi:10.3200/JOER.99.2.67-77

Jacobson, J., Thrope, L., Fisher, D., Lapp, D., Frey, N., & Flood, J. (2001). Cross-age tutoring: A literacy improvement approach for struggling adolescent readers. *Journal of Adolescent & Adult Literacy, 44,* 528–536.

Kieran, C. (1992). The learning and teaching of school algebra. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 390–419). New York, NY: MacMillan.

King, A. (1998). Transactive peer tutoring: Distributing cognition and metacognition. *Educational Psychology Review, 10,* 57–74. doi:10.1023/A:1022858115001

King, A., Staffieri, A., & Adelgais, A. (1998). Mutual peer tutoring: Effects of structuring tutorial interaction to scaffold peer learning. *Journal of Educational Psychology, 90,* 134–152. doi:10.1037/0022-0663.90.1.134

Lau, T. A., & Weld, D. S. (1998). Programming by demonstration: An inductive learning formulation *Proceedings of the 4th International Conference on Intelligent User Interfaces* (pp. 145–152). New York, NY: ACM Press.

Leelawong, K., & Biswas, G. (2008). Designing learning by teaching agents: The Betty's Brain system. *International Journal of Artificial Intelligence in Education, 18,* 181–208.

Li, N., Matsuda, N., Cohen, W. W., & Koedinger, K. R. (2011). A machine learning approach for automatic student model discovery. In M. Pechenizkiy, T. Calders, C. Conati, S. Ventura, C. Romero, & J. Stamper (Eds.), *Proceedings of the International Conference on Educational*

Data Mining (pp. 31–40). Eindhoven, the Netherlands: International Educational Data Mining Society.

Linchevski, L., & Herscovics, N. (1996). Crossing the cognitive gap between arithmetic and algebra: Operating on the unknown in the context of equations. *Educational Studies in Mathematics, 30,* 39–65. doi:10.1007/BF00163752

Mastropieri, M. A., Spencer, V. G., Scruggs, T. E., & Talbott, E. (2000). Students with disabilities as tutors: An updated research synthesis. In T. E. Scruggs & M. A. Mastropieri (Eds.), *Advances in learning and behavioral disabilities* (Vol. 14, pp. 247–279). Stamford, CT: JAI Press.

Mathan, S. A., & Koedinger, K. R. (2005). Fostering the intelligent novice: Learning from errors with metacognitive tutoring. *Educational Psychologist, 40,* 257–265. doi:10.1207/s15326985ep4004_7

Mathes, P. G., & Fuchs, L. S. (1994). The efficacy of peer tutoring in reading for students with mild disabilities: A best-evidence synthesis. *School Psychology Review, 23,* 59–80.

Matsuda, N., Cohen, W. W., Koedinger, K. R., Keiser, V., Raizada, R., Yarzebinski, E., & Stylianides, G. J. (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., Cohen, W. W., Sewall, J., Lacerda, G., & Koedinger, K. R. (2007). Evaluating a simulated student using real students data for training and testing. In C. Conati, K. McCoy, & G. Paliouras (Eds.), *Proceedings of the International Conference on User Modeling (LNAI 4511)* (pp. 107–116). Berlin, Germany: Springer.

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. J., Cohen, W. W., & Koedinger, K. R. (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). Berlin, Germany: Springer.

Matsuda, N., Yarzebinski, E., Keiser, V., Raizada, R., Stylianides, G. J., & 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 the International Conference on Intelligent Tutoring Systems* (pp. 101–111). Heidelberg, Germany: Springer-Verlag.

Michie, D., Paterson, A., & Hayes, J. E. (1989). *Learning by teaching. Proceedings of the Second Scandinabian Conference on Artificial Intelligence* (pp. 413–436). Amsterdam, the Netherlands: IOS.

Ogan, A., Finkelstein, S., Mayfield, E., D'Adamo, C., Matsuda, N., & Cassell, J. (2012). "Oh, dear Stacy!" Social interaction, elaboration, and learning with teachable agents. In *Proceedings of the International Conference on Computer–Human Interaction (CHI2012)* (pp. 39–48). Austin, TX: ACM.

Ohlsson, S. (2008). Computational models of skill acquisition. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 359–395). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511816772.017

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1,* 117–175. doi:10.1207/s1532690xci0102_1

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.), *Proceedings of the International Conference on Artificial Intelligence in Education* (pp. 247–255). Heidelberg, Germany: Springer.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1,* 81–106. doi:10.1007/BF00116251

Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review, 14,* 249–255. doi:10.3758/BF03194060

Robinson, D., Schofield, J., & Steers-Wentzell, K. (2005). Peer and cross-age tutoring in math: Outcomes and their design implications. *Educational Psychology Review, 17,* 327–362. doi:10.1007/s10648-005-8137-2

Rohrbeck, C. A., Ginsburg-Block, M., Fantuzzo, J. W., & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology, 95,* 240–257. doi:10.1037/0022-0663.95.2.240

Roscoe, R. D., & Chi, M. T. H. (2004). The influence of the tutee in learning by peer tutoring. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (pp. 1179–1184). Mahwah, NJ: Erlbaum.

Roscoe, R. D., & Chi, M. T. H. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research, 77,* 534–574. doi:10.3102/0034654307309920

Sharpley, A. M., Irvine, J. W., & Sharpley, C. F. (1983). An examination of the effectiveness of a cross-age tutoring program in mathematics for elementary school children. *American Educational Research Journal, 20,* 103–111.

Topping, K. J. (1996). The effectiveness of peer tutoring in further and higher education: A typology and review of the literature. *Higher Education, 32,* 321–345. doi:10.1007/BF00138870

Ur, S., & VanLehn, K. (1995). STEPS: A Simulated, Tutorable Physics Student. *Journal of Artificial Intelligence in Education, 6,* 405–437.

VanLehn, K., Ohlsson, S., & Nason, R. (1994). Applications of simulated students: An exploration. *Journal of Artificial Intelligence in Education, 5,* 135–175.

Vassileva, J., McCalla, G., & Greer, J. (2003). Multi-agent multi-user modeling in I-Help. *User Modeling and User-Adapted Interaction, 13,* 179–210. doi:10.1023/A:1024072706526

Walker, E., Rummel, N., & Koedinger, K. R. (2009). CTRL: A research framework for providing adaptive collaborative learning support. *User Modeling and User-Adapted Interaction, 19,* 387–431. doi:10.1007/s11257-009-9069-1

Webb, N. M., & Mastergeorge, A. (2003). Promoting effective helping behavior in peer-directed groups. *International Journal of Educational Research, 39,* 73–97. doi:10.1016/S0883-0355(03)00074-0

Received December 16, 2011
Revision received November 16, 2012
Accepted December 10, 2012 ■