RESEARCH ARTICLE

Studying the Effect of a Competitive Game Show in a Learning by Teaching Environment

**Noboru Matsuda · Evelyn Yarzebinski · Victoria Keiser ·
Rohan Raizada · Gabriel J. Stylianides · Kenneth R. Koedinger**

Abstract In this paper we investigate how competition among tutees in the context of learning by teaching affects tutors' engagement as well as tutor learning. We conducted this investigation by incorporating a competitive Game Show feature into an online learning environment where students learn to solve algebraic equations by teaching a synthetic peer, called SimStudent. In the Game Show, pairs of SimStudents trained by students beforehand competed against each other by solving challenging problems to attain higher ratings. The results of a classroom study with 141 7th through 9th grade students showed the following: (1) Students improved their proficiency to solve equations after teaching SimStudent, but there was no observed improvement in their conceptual understanding. (2) Overall, the competitive Game Show promoted students' extrinsic and intrinsic motivations—when the competitive Game Show was available, students'

N. Matsuda (✉) · E. Yarzebinski · K. R. Koedinger
Human-Computer Interaction Institute, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA
e-mail: Noboru.Matsuda@cs.cmu.edu

E. Yarzebinski
e-mail: eey2@cs.cmu.edu

K. R. Koedinger
e-mail: Koedinger@cs.cmu.edu

V. Keiser
Carnegie Learning Inc., 437 Grant Street, Pittsburgh, PA 15219, USA
e-mail: victoriakeiser@gmail.com

R. Raizada
Akamai Technologies India Pvt Ltd., "A" Wing, Survey No.80/1, Bellandur Village Outer Ring Road, Bangalore 560037, India
e-mail: rraizada@akamai.com

G. J. Stylianides
Department of Education, University of Oxford, 15 Norham Gardens, Oxford OX2 6PY, UK
e-mail: gabriel.stylianides@education.ox.ac.uk

engagement in tutoring (intrinsic motivation) was increased; students who arguably had a higher desire to win strategically selected opponents with lower proficiency for an easy win (extrinsic motivation). (3) The availability of the competitive Game Show did not affect tutor learning; there was no notable correlation between students' motivation (intrinsic or extrinsic) and tutor learning. Based on these findings, we propose design improvements to increase tutor learning.

Keywords Learning by teaching · Teachable agent · Motivation and engagement · SimStudent · Algebra

Introduction

The goal of this paper is to investigate how adding a competitive Game Show in the context of learning by teaching affects students' engagement in tutoring their synthetic peers, and whether increased engagement (if any) facilitates tutor learning. The present study is a part of our on-going effort (Matsuda et al. 2012a, b, 2011, 2013) to understand the cognitive and social factors that govern learning by teaching (Cohen 1994; Cohen et al. 1982; Hedin 1987; Roscoe and Chi 2007).

Despite the long history of research on tutor learning (Chi et al. 2001; Cohen et al. 1982; Devin-Sheehan et al. 1976; Gartner et al. 1971; Graesser et al. 1995), it has been only recently that researchers have started to explore the underlying principles of tutor learning. This intellectual evolution is largely due to the recent developments of educational technology to build teachable agents.

Using a teachable agent, researchers can collect detailed *process data* documenting interactions between students and the teachable agent. Collecting process data from field studies where human students teach each other is very challenging and seldom done (Roscoe and Chi 2007). When combined with the learning *outcome data* that are typically test scores, process data allow researchers to examine research questions about how learning unfolds and how affordances of the instructional environment change learning. To address this issue, we have developed an online learning environment in which students learn by teaching a synthetic peer, called SimStudent. In this particular paper, our primary focus is on the motivational factors and their impact on tutor learning (Matsuda et al. 2012b).

SimStudent is a machine-learning architecture that learns procedural skills inductively from examples (Matsuda et al. 2007, 2008, 2009). One of the applications of SimStudent is to use it as a teachable agent (VanLehn et al. 1994). When integrated into the online learning environment, called APLUS (Artificial Peer Learning environment Using SimStudent), students can interactively tutor SimStudent.

In our previous studies with SimStudent, we have observed that students interacting with SimStudent often express excitement. In the current study, we introduced the competitive Game Show to further stimulate students' motivation. The students' goal is to tutor their own SimStudent to participate in the Game Show and to earn the highest rating possible, which is determined based on the results of competitions. In an individual competition, two SimStudent agents compete against each other by solving problems entered by students and the Game Show host.

We believe that the more the students care about the proficiency of their synthetic peers the more deeply students will engage in tutoring them, which in turn can facilitate students' learning (i.e., tutor learning). More specifically, we hypothesize that the competitive Game Show will promote students' engagement in tutoring their peers (intrinsic motivation), and the deeper engagement will in turn facilitate tutor learning. We also hypothesize that the competitive Game Show will increase students' desire to win the competition (extrinsic motivation).

There have been studies on how a competitive task affects intrinsic motivation, but the results are mixed. Some studies showed that competition negatively affected intrinsic motivation (Deci et al. 1981; Reeve and Deci 1996; Reeve et al. 1985) while some showed positive influence (Chase et al. 2009; Vansteenkiste and Deci 2003). Therefore, it is not clear how our Game Show intervention will affect tutor learning. To address this issue, we conducted a classroom-based, tightly controlled in vivo experiment (cf., Koedinger et al. 2009) to study the effect of APLUS with and without the Game Show.

In the rest of the paper, we first discuss the theoretical background of our research questions and hypotheses. We then provide a technical overview of SimStudent, APLUS, and the competitive Game Show. We describe the classroom study in which the effect of the competitive Game Show was tested. Finally, we present the results of the classroom study followed by a discussion of implications, limitations, and directions for future research.

Competitive Game Show in Learning by Teaching

There are two primary reasons that inspired us to hypothesize that the Game Show would promote students' engagement in tutoring (intrinsic motivation), but at the same time, it would also promote students' desire to win the game (extrinsic motivation).

First, although the competitive Game Show provides competitively-contingent reward that is potentially harmful for tutor learning (Vansteenkiste and Deci 2003), it might promote students' intrinsic motivation. This is because it is SimStudents who compete, not students themselves. Students with this perception might still be intrinsically motivated to tutor their SimStudents even when students are competitively motivated. The protégé effect, i.e., treating the synthetic peer as the agent to be judged, is known to have a notable effect for tutor learning (Chase et al. 2009).

Second, students' intrinsic and extrinsic motivation would also be reflected in the strategy to win the competition. Some students might choose to make their SimStudent stronger by tutoring difficult problems, which would positively facilitate tutor learning, whereas some might choose to obtain an easy win by strategically selecting lower rated opponents, which might be detrimental for tutor learning.

Based on these observations, we operationalize intrinsic and extrinsic motivations in the current study as follows. We operationalize *intrinsic motivation* behaviorally (Deci 1971; Reeve and Deci 1996) as a students' commitment to tutoring as measured by engagement factors that we discuss in the "Measures" section. We operationalize *extrinsic motivation* as the inclination toward winning the game.

In sum, we investigate the following research questions: (Q1) Do students learn by teaching a synthetic peer? (Q2) How does the competitive Game Show affect

students' tutoring engagement and desire to win? (Q3) How do engagement in tutoring and desire to win affect tutor learning?

To address these research questions, we tested the following hypotheses, each corresponding to a research question: (H1) The proposed online system would facilitate tutor learning both in procedural and conceptual knowledge. (H2) The competitive Game Show would increase students' motivation for not only winning but also for tutoring SimStudent better. At the operational level, we should see an increase in their engagement for tutoring in terms of more problems tutored, more responses to SimStudent's questions, more quizzes completed, etc. (H3) The increased tutoring engagement would positively affect tutor learning whereas the increased competitively-oriented motivation would positively or negatively affect tutor learning depending on the strategy to win the competition. For the latter, there would be two primary strategies: (a) to tutor SimStudent hard to make it more proficient that requires better tutoring, which should positively affect tutor learning, and (b) to compete against lower rated opponents for an easy win, which should negatively affect tutor learning.

Learning Environment: SimStudent, APLUS, and the Game Show

This section provides a brief overview of SimStudent, APLUS, and Game Show. Figure 1 shows an annotated screenshot as a reference for descriptions below.

Overview of SimStudent

In the context of learning by teaching, SimStudent learns cognitive skills through tutored problem-solving, in which students interactively tutor SimStudent (Matsuda et al. 2012a, b, 2010, 2011). A student poses a problem for SimStudent to solve. SimStudent attempts to solve the problem step by step by suggesting one step at a time through application of its existing knowledge. SimStudent then asks about the correctness of the suggestion, and the student provides yes/no feedback (often called flagged feedback). If the feedback is negative, then SimStudent may suggest an alternative action on the step. When SimStudent has no other suggestions, it asks the student what to do next. The student then demonstrates the next step as a hint for SimStudent.

SimStudent learns skills by inductively generalizing examples provided by the student. Positive feedback and steps demonstrated by the student as responses for hint requests become *positive examples*. Negative feedback becomes a *negative example*. Using the technique of programming by demonstration (Lau and Weld 1998), SimStudent generalizes positive and negative examples and generates a set of hypotheses that explain examples in the form of production rules. Each production models a single skill representing *where* to pay attention (among the interface elements) to determine *when* (in terms of the conditions held among the focus of attention) and *how* to make a particular step (as a result of applying a chain of operations). SimStudent applies three different Artificial Intelligence techniques to learn each part of the production—the version space (Mitchell 1997) for the where-part, the inductive logic programming (Muggleton and de Raedt 1994) implemented as FOIL (Quinlan 1990) for the when-
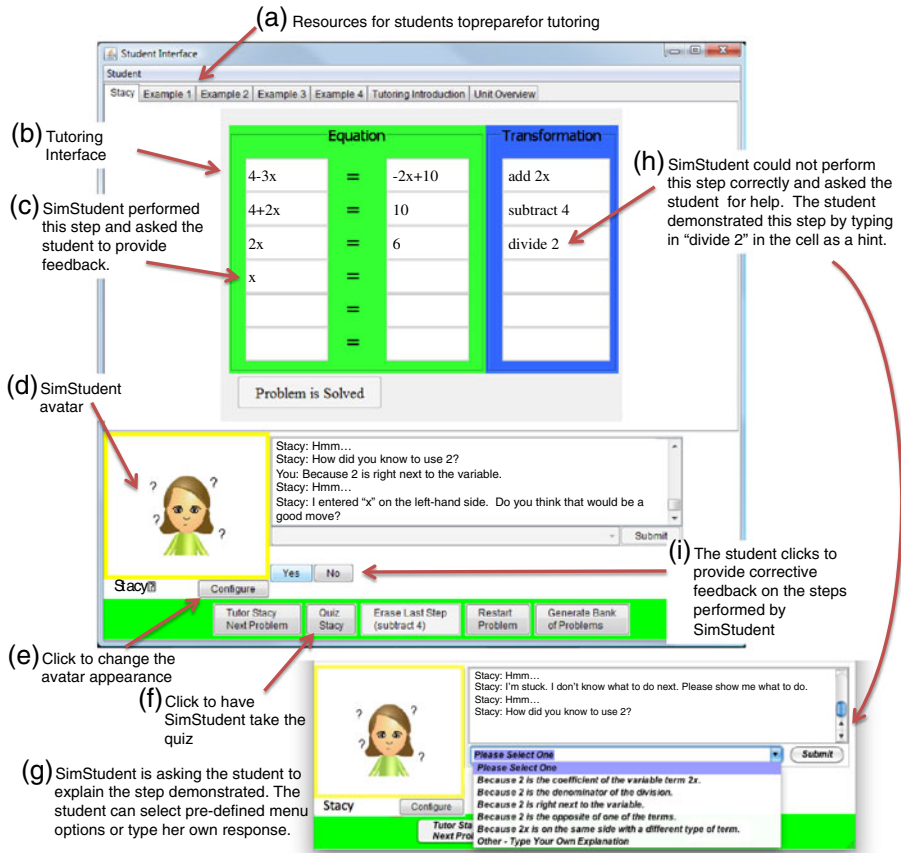
Fig. 1 An example screenshot of APLUS

part, and the iterative-deepening depth-first search (Russell and Norvig 2003) for the how-part. Providing technical details of the learning algorithm is beyond the scope of the current paper, but can be found elsewhere (Matsuda et al. 2007, 2008).

Because of its genuine machine-learning capability, SimStudent learns skills differently depending on how students tutor SimStudent. Students often tutor SimStudent incorrectly and inappropriately (Matsuda et al. 2013). For example, in the situation depicted in Fig. 1-h, the student could incorrectly demonstrate "divide 6," which is a common mistake that many high school students make. In such a situation, SimStudent might learn to "divide both sides by the constant on the right hand side," which would suggest for SimStudent to, say, "divide 2" for "5x=2."

One of the unique characteristics of SimStudent as a teachable agent is its ability to model both correct *and* incorrect learning. We have demonstrated that when SimStudent is given prior knowledge that relies on surface features without a connection to the deep domain semantics, it makes induction errors from correct examples. In the context of learning by teaching, even when students tutor skills correctly the incorrect induction results in learning overly specific or overly general productions that represent errors that students commonly make in algebra (Booth and Koedinger 2008; Matsuda et al. 2009).

The cognitive fidelity of learning, especially incorrect learning, is one of the strengths of SimStudent as a teachable agent. We hypothesize that teachable agents with high cognitive fidelity would provide better opportunities for students to learn by teaching. There are other types of teachable agents that do not commit to actual learning, i.e., there is no machine-learning technique involved to interactively and dynamically acquire new skills. For example, some teachable agents (e.g., Pareto et al. 2011) use the knowledge-tracing technique (Corbett and Anderson 1995) in which all necessary knowledge is given a priori. Others merely interpret shared problem-solving skills (Biswas et al. 2005; Bredeweg et al. 2009).

Overview of APLUS

SimStudent is visualized with an avatar at the bottom left corner of the APLUS interface (Fig. 1-d). Students can customize the avatar image (i.e., eyes, nose, hair, color, etc.) and name it, e.g., Stacy, as shown in the figure. Students can customize the avatar anytime by clicking on the [Configure] button (Fig. 1-e).

Students tutor SimStudent using the Tutoring Interface shown in Fig. 1-b. To pose a new problem for SimStudent, students enter an equation into the first row of the Tutoring Interface. The steps performed by SimStudent are displayed in the Tutoring Interface one step at a time (Fig. 1-c). Students provide feedback on each step performed by SimStudent using the [Yes/No] button (Fig. 1-i). When SimStudent cannot make a "correct" attempt, it asks students for a hint. Students then must demonstrate the next step in the Tutoring Interface (Fig. 1-h).

SimStudent occasionally asks questions to prompt students to explain their tutoring decisions (Matsuda et al. 2012a) in the following situations: (1) When the student poses a new problem, SimStudent asks the reason for selecting a particular problem to solve. (2) When the student provides negative feedback on the step that SimStudent performed, SimStudent asks how the situation is different from a previous step on which the same skill was used and where positive feedback was received (e.g., "But I put 'divide 3' for $3x=9$. Why doesn't 'divide 2' work now?"). (3) When the student provides a hint about a transformation step, SimStudent asks the students to explain the step.

The student responds to SimStudent's question either by free text input or using drop down menu, depending on the type of question. For questions about negative feedback, students enter their own answers as free text. For a question on a new problem or hint, a drop down menu is available. Students can also edit the selected menu item to add their own words or completely re-write the response. Figure 1-g shows an example of SimStudent asking a question about a step demonstrated by the student as a hint and the drop down menu. SimStudent does not understand student's responses, whether free text or menu selection; the question answering was designed to promote students' reflective thinking that the theory of self-explanation predicts as an effective learning strategy (Aleven and Koedinger 2002; Chi 2000).

Students can quiz SimStudent to assess its competency anytime during tutoring by clicking on the [Quiz] button (Fig. 1-f). The quiz has four sections with two equations each. Section 1 has a one-step equation (e.g., $x+5=10$) and a two-step equation (e.g., $2x-3=7$), section 2 has two two-step equations, and sections 3 and 4 have two equations each with variables on both sides (e.g., $3x+1=5-2x$). SimStudent takes

the quiz one section at a time and must solve both equations correctly to proceed to the next section. The equations are randomly generated each time SimStudent takes the quiz, but the *type* of equation is fixed—i.e., only constants and coefficients of equations are randomly generated each time SimStudent takes the quiz. After SimStudent takes the quiz, the summary of the results are displayed with the correctness of the steps being graded by the embedded Cognitive Tutor Algebra (Ritter et al. 2007). In the summary, correct steps are displayed in green and incorrect steps are displayed in red.

If students get stuck, APLUS provides several resources to assist students tutoring SimStudent (Fig. 1-a). Clicking on the [Example] tabs shows worked-out examples in the Tutoring Interface with annotated explanations. The Unit Overview provides descriptions of the process of solving equations. The Tutoring Introduction is a short video showing how to use APLUS. Before students started tutoring, they watched the introductory video together on a projector screen in front of the classroom. Students also had the option of watching the video again individually during the study, if they felt they needed to do so.

Overview of the Game Show

Figure 2 shows a sample screenshot of the Game Show. The Game Show is displayed on each student's screen as a separate window from APLUS. Two SimStudent avatars (Stacy and Amy in Fig. 2) are displayed in the middle of the window with a Game Show host on the left.

In a single Game Show session, SimStudents solve five problems. The Game Show host provides the first problem that is randomly selected from 40 pre-specified kinds of problem with randomly generated constants and coefficients. The two student contestants (i.e., the students who tutored the competing SimStudents) then take turns providing two problems each.

When SimStudents solve problems, students see their own SimStudent working and filling in the problem-solving interface—same as the Tutoring Interface on APLUS—on their own Game Show window. When SimStudents solve all five
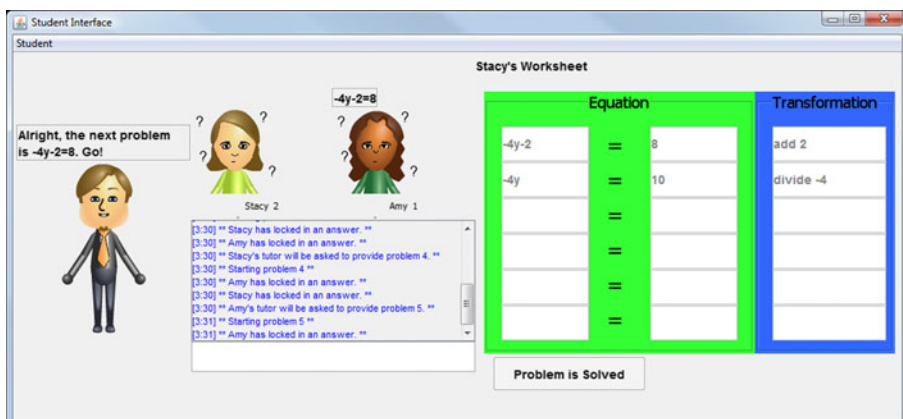

Fig. 2 The Game Show window. A pair of SimStudents compete by solving five problems entered by the students and the game-show host, which is displayed on the left corner

problems, the Game Show host brings up a review screen in which the students can review and compare the solutions that two SimStudents made for each problem. The correctness of each step is indicated in a different color. Our expectation was that by reviewing the work of their SimStudent (and comparing different solutions), students would know the weaknesses of their SimStudent and would, therefore, be motivated to continue tutoring.

In the Game Show, SimStudents are rated. All SimStudents start at a rating of 25. Ratings are calculated based on the relative rating of the SimStudent who won and the one who lost (computed once after all five problems are solved). The calculation is similar to the ELO chess rating system (Elo 1978)—the winner's rating increases, and the loser's rating decreases. The amount of gain (or loss) is proportional to the difference in the ratings between the two contestants. For example, when winning against a higher-rated opponent, the bigger the difference between opponents, the bigger the gain, whereas when losing against a lower-rated opponent, the bigger the difference between opponents, the bigger the loss. Even when one wins against a lower rated opponent, the winner's score increases, even if by just a small amount.

Before entering the Game Show, students select their opponents on the match-up screen as shown in Fig. 3. In the match-up screen, a list of students waiting to be matched-up is shown. They can also chat with each other. When a student selects an opponent, the opponent's SimStudent avatar is displayed along with the profile of the opponent showing its rating and history of game results. As far as we know, students tended not to use their names for their SimStudent's name. Although some students attempted to identify who was the tutor of a particular SimStudent via chat (e.g., "who is dyalan") there was no evidence in the chat logs in the process data to show that students' identities were actually revealed. Thus, it would be fair to say that each match-up was totally anonymous.

The expected rating after a win or loss is also shown on the screen. Students can challenge any students on the waiting list, or they can wait for someone to issue a challenge to them. When receiving a challenge, students will see the challenger's
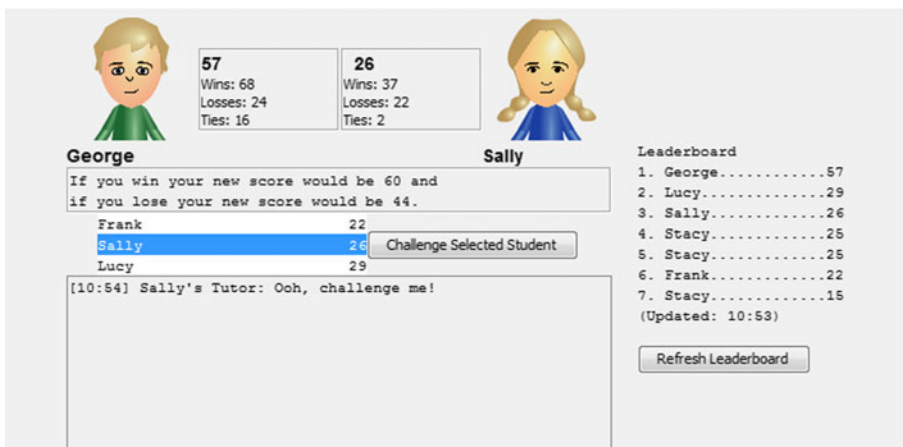
Fig. 3 The match-up screen of the Game Show. The student can select an opponent from the list. A profile for the selected opponent is then displayed with his/her rating and the history of game results. Students can also chat with each other

profile and expected rating after the game. The student being challenged must immediately accept or reject another students' request for a game.

Evaluation Study

Structure of the Study

The study was a randomized control trial with the experimental condition using APLUS with the Game Show, and the control condition using APLUS only. Seven classes from one high school in Pittsburgh, PA, participated in the study under the supervision of the Pittsburgh Science of Learning Center (www.learnlab.org).

The study was conducted as a class-level randomization among six classes and a within class randomization for one class. We chose a class-level randomization, because the difference between two conditions is quite discernible. That is, since the appearance of the Game Show was arguably attractive, there was a potential threat that students in the non-Game Show condition might be demotivated as we observed in a previous study in which we compared SimStudent (that was novel to the participants) and the Cognitive Tutor Algebra (that participants used regularly in the algebra class).

The students in both conditions were told that their task was to teach SimStudent to solve equations with variables on both sides. Students in the control condition were told that their goal was to have SimStudent pass the quiz. Students in the experimental condition were told that their goal was to attain the highest rating in the Game Show. Students in the experimental condition were also told that they could switch between APLUS and the Game Show as often as they liked.

The intervention lasted during three regular class periods (a single class period lasted for 42 min) on three consecutive days. There was a pre- and post-test on the days immediately before and after the intervention periods, and a delayed test 2 weeks after the post-test.

Measures

Students' learning was measured and analyzed using the outcome and process data. The outcome data were measured using online tests, whereas the process data were collected automatically during the intervention.

Students' learning gain was measured with an online test that consists of two parts—a procedural skill test and a conceptual knowledge test. Three isomorphic versions of the test were created. To counterbalance any potential test differences, each individual student was assigned a random ordering of the three versions of the test for pre-, post-, and delayed-tests. The procedural skill test had three sections: (a) ten equation solving items, (b) twelve items to determine if a given operation is a logical next step for a given equation, and (c) five items to identify the incorrect step in a given incorrect solution. The conceptual knowledge test had two sections divided into: (d) 38 true/false items about basic algebra vocabulary, and (e) ten true/false items to determine if two given algebra expressions are equivalent. To increase the reliability of the measures, we aggregate items into two tests based on higher-level skill type (as opposed to treating each item as an individual measure) in the following analysis.

The process data showed students' interactions with the system (APLUS, SimStudent, and Game Show) including problems tutored, feedback provided, steps performed, examples reviewed, hints requested by SimStudent, quiz attempts, Game Show participation, Game Show wins and losses, Game Show rating, and Game Show opponents challenged.

To quantify students' engagement during tutoring, several variables in the process data that might have reflected the degree of students' commitment and care about their SimStudents' learning were used. One such example is the quality of self-explanation students provided. SimStudent asked occasional "why" questions 3,938 times with the average number of questions per student greater for Baseline (BL) students than Game Show (GS) students; $M_{BL}=53.1$ ($SD=16.3$) vs. $M_{GS}=36.4$ ($SD=15.2$); $t(86)=5.00$, $p<0.001$. As described below, since Game Show students spent less time on tutoring to participate to the Game Show when the total study time was controlled, this difference was not surprising. For those 3,938 SimStudents' questions, students actually responded 3,652 times with the average per student $M_{BL}=48.3$ ($SD=5.2$) vs. $M_{GS}=34.7$ ($SD=14.5$).

Three human coders categorized these 3,652 responses into "deep" and "shallow" responses. We computed the inter-coder reliability using Cohen's Kappa. However, there was an unfortunate data loss including the Kappa value. Thus, we can no longer report the inter-coder reliability value. We have the inter-coder reliability value for pilot coding that was computed when we developed the coding manual for this analysis. The Kappa value ranged from 0.62 to 0.77.

Following the post-test, students were invited to take a questionnaire. The questionnaire is a modified version of the contextualized measures of achievement goals, self-efficacy, and affect (Bernacki et al. 2013). The questionnaire is comprised of (a) 16 items on a 7-point Likert-scale that measured different types of motivation and (b) one free response item on the ease or difficulty of tutoring SimStudent. There are four constructs on the questionnaire with reliabilities of 0.79 (mastery), 0.77 (performance), 0.55 (strategy), and 0.49 (affect). Because the affect construct reliability was low, we excluded the questionnaire items about affect from the analysis.

Participants

There were a total of 141 students about evenly placed in each of the seven classes. Of those 141 students, 106 (75.2 %) students took the pre-test with exactly 53 students in each condition. Of those 106 students, 88 (83.0 %) students participated in all three intervention days and took the post-test: again there were equal numbers of students in each condition. Of those 88 students, 69 (78.4 %) students also took the delayed-test: there were 29 and 40 students in the study and the control conditions respectively who took the delayed-test. Of those 69 students, 59 students took the questionnaire.

It is often observed that the attrition is more or less impacted by low achieved students who generally have low motivation in completing a classroom experiment. This bias often affects the results of the study as well—higher post (or delayed) test scores in one condition might be simply due to the exclusion of lower motivated students who dropped from the study. This could be troublesome in our analysis, because the number of students who took the delayed test was different between the

two conditions. In our study, however, there were no notable condition differences on pre-test ($t(4.51)=-1.43$, $p=0.22$) nor post-test ($t(5.56)=0.56$, $p=0.60$) scores among those 19 students who did not take the delayed-test, suggesting that we can safely eliminate the potential threat of the motivation bias as a cause of the condition difference in delayed-test scores (if any).

The following analyses include those students who took all three tests and participated during all 3 days of the study (i.e., $N=69$, unless otherwise specified).

Results

Test Scores

Did students learn by teaching SimStudent? If so, did Game Show students learn more than baseline students? To answer this question, we ran a mixed-design analysis with test-time (pre, post, and delayed) as a within-subject variable and condition as a between-subject variable. The analysis was run for the procedural skill test and the conceptual knowledge test separately. Figure 4 shows the test scores.

For the procedural skill test, the average scores of both post-test ($M_{Post}=.45$, $SD=.22$) and delayed-post test ($M_{Delayed}=.46$, $SD=.23$) were significantly higher than the pre-test ($M_{Pre}=.38$, $SD=.20$); $t(68)=-3.61$, $p<0.001$ and $t(68)=-4.31$, $p<0.001$ for pre vs. post and pre vs. delayed respectively. The difference between the post- and delayed-test was not statistically significant.

There was a marginal difference between the conditions suggesting that Game Show (GS) students performed slightly better on the three tests than Baseline (BL) students. Unfortunately, we had a non-ideal randomization such that there were significant differences on the pre-test: $M_{GS}=.44$ ($SD=.19$) vs. $M_{BL}=.34$ ($SD=.19$); $t(60.3)=-2.14$, $p<0.05$. When post-test scores are adjusted for pre-test, $M_{GS}=.48$($SD=0.12$) vs. $M_{BL}=.42$
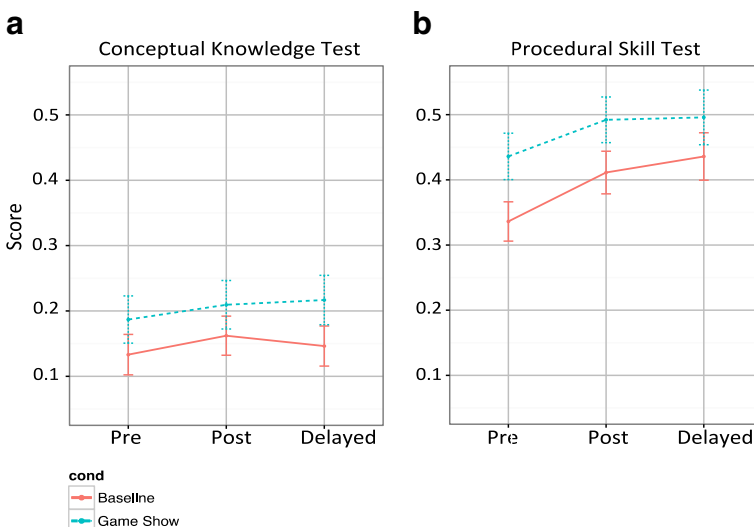
Fig. 4 Test scores for the procedural skill test (**a**) and the conceptual knowledge test (**b**)

(*SD*=0.12), there is a trend in favor of the Game Show group. However, an ANCOVA using the pre-test score as a covariate did not confirm a statistically reliable condition difference on the post-test of the procedural skill test; $F(1,66)=0.23$, $p=0.63$.

As for the conceptual knowledge test, there was no test score difference among test-time (pre, post, and delayed) or between conditions.

Motivation and Tutoring Engagement

Why was there no significant condition difference in students' learning? To answer this question, we investigated whether Game Show students were actually more motivated in tutoring SimStudent than Baseline students, and if so how they were engaged in tutoring. In this analysis we used the questionnaire to quantify the motivation and process data to measure the engagement as mentioned in section "Measures."

There was no condition difference in the mean response on any of three question-naire constructs. According to the students' self-reports, students in the two conditions were equally concerned with mastering the subject matter, they were equally strategic in tutoring SimStudent, and they expressed equal affective status. There was no notable correlation between the motivation factors and the learning gain on the test scores (both on the procedural skill test and the conceptual knowledge test).

What about the actual engagement in tutoring? Was there any condition difference in the way students tutored SimStudent? We analyzed how students tutored SimStudent using the engagement variables mentioned in the section "Measures." It turned out that the Game Show (GS) students responded to SimStudent's questions more often than the Baseline (BL) students; the mean probability of responding to SimStudent's questions was $M_{GS}=.93$ (*SD*=0.12) vs. $M_{BL}=.83$ (*SD*=0.25); $t(63)=2.27$, $p<.05$. The higher probability of answering SimStudent's questions indicates that the Game Show students might have been more motivated to tutor SimStudent more seriously.

The Game Show students were also more likely to enter a "deep" response to SimStudent's questions than the Baseline students; the mean probability of entering a "deep" response was $M_{GS}=.24$ (*SD*=0.18) vs. $M_{BL}=.16$ (*SD*=0.15); $t(85)=-2.34$, $p<.05$. On the other hand, students in the two conditions used an equal amount of words per response—suggesting that the Game Show students were more thoughtful in entering responses than the Baseline students who were equally talkative, but they tended to enter "shallow" responses.

The Game Show students tutored significantly fewer problems than the Baseline students; the mean number of tutored problems was $M_{GS}=11.1$ (*SD*=5.0) vs. $M_{BL}=18.6$ (*SD*=7.57); $t(75)=5.45$, $p<.001$. Game Show students spent significantly less time on tutoring (problem completed and abandoned) than baseline students; $M_{GS}=34.8$ min (*SD*=17.8) vs. $M_{BL}=49.7$ min (*SD*=16.2), $t(86)=4.10$, $p<0.001$. There was no difference in the time spent on each problem to tutor; $M_{GS}=1.7$ min (*SD*=0.64) vs. $M_{BL}=1.8$ min (*SD*=0.68), $t(84)=-0.67$, p=0.50.

There were a few statistically significant correlations between motivation (i.e., questionnaire response) and engagement (i.e., actual tutoring activities recorded in the process data), but their correlation coefficients were relatively small. There was no significant correlation between motivation and learning gain measures (both on the procedural skill test and the conceptual knowledge test).

These results imply that the Game Show students were more engaged in tutoring than the Baseline students while students in both conditions self-reported the same motivation levels. The Game Show students seemed to have stronger desires to have their SimStudent learn how to solve equations than the Baseline students. Yet, there was no condition difference for the students' learning. To further investigate how students participated in the Game Show, we analyzed the process data as shown in the next section.

Game Show Participation

How did students participate in the Game Show? Was a particular strategy for participation helpful or detrimental for tutor learning? To answer these questions, we first quantified students' participation in the Game Show. Table 1 shows descriptive statistics of the Game Show participation. Since we only analyze Game Show students, the data analysis in this section was conducted with 44 students in the Game Show condition who took both pre- and post-test to gain more statistical power.

The data show that students were actively involved in the Game Show: students spent about 36 % of the study time on the Game Show. On average, students challenged 4.5 times more than they accepted a challenge from others. The probability of tutoring SimStudent after a win was very low, which is not surprising. However, the probability of returning to tutoring SimStudent when students lost the game was also quite low—showing that *once students entered the Game Show, they tended to remain in the Game Show instead of tutoring SimStudent to make it stronger.*

To understand how alternating Game Show and tutoring affected students' learning, we analyzed the tendency of tutoring after Game Show as the ratio of tutoring after a competition (TAC). We then compared the average TAC ratio after a win and loss, and found no difference; M_{win}=.12 (SD=.19) vs. M_{loss}=.21 (SD=.32); $t(27)$=−1.283, p=0.21 (there were only 28 students who had both wins *and* losses). There was also no difference in TAC (aggregated across wins and losses) between the 1st and the 4th quartiles of the normalized gain for the procedural skill test: M_{1st}=.15 (SD=.22) vs. M_{4th}=.21 (SD=.27); $t(19)$=−.590, p=0.56. In sum, students' tendencies in tutoring after a competition was not predictive of their gain on the procedural skill test regardless of the results of the competition, though this is arguably due to the low TAC ratio.

The average number of competition per student was 7.5 (SD=4.5). There was no competition observed on Day 1, but there were a total of 408 competitions on Days 2 and 3, with 2.1 times as many competitions held on Day 3 than Day 2. This implies

Table 1 Descriptive statistics of the Game Show participation. The table shows a mean±SD of each variable

Variable	M±SD	Variable	M±SD
Time on the Game Show	45.3±26.4 min	Ratio of challenge/accept	4.5±3.6
Probability of tutoring after win	.12±0.19	Probability of tutoring after loss	.21±0.32
Challenging delta	−2.4±13.8	Accepting delta	−2.0±11.5
Final rating	26.8±16.0		

that students did not rush to play the Game Show, but rather spent time tutoring on Day 1 and some on Day 2 to prepare for the competition. Together with the data showing students' engagement mentioned above, *students were arguably considerate about their SimStudent's proficiency to enter the Game Show.*

The average final rating of 26.8±16.0 is surprisingly low given that the initial rate was 25. Most surprisingly, *students tended to challenge against and accept a challenge from lower rated students*—the average challenging delta, which is the average rating difference when challenging, was −2.4 (MIN=−56, MAX=60) and the average accepting delta, which is the average rating difference when accepting a challenge, was −2.0 (MIN=−60, MAX=56). For both variables, the difference is relative to the opponents' rating—i.e., opponent's rating minus the student's rating.

Overall, the above findings imply that *students' primary goal was to attain higher rating that motivated students to strategically select an easy win rather than training SimStudent better (which would have also resulted in a higher rating).* This indicates that students' performance goal was not aligned with the learning goal, seen also through there being no notable correlation between the final rating and the normalized gain from pre- to post- procedural skill test.

How do engagement in tutoring and desire to win affect tutor learning? To answer this question, we grouped the students based on their preference in selecting opponents. Figure 5 shows a scatter plot for the average difference of rating when students challenged, i.e., the *challenging delta* (X-axis) and accepted other's requests, i.e., the *accepting delta* (Y-axis).

There was a strong correlation between the challenging and accepting deltas; $r(40)=0.57$, $p<0.01$. Those who tended to challenge higher (lower) rated students also tended to accept challenges from higher (lower) rated students.

In Fig. 5, the top right quadrant shows students who challenged higher rated opponents and accepted challenge requests from higher rated opponents. This group of students could be labeled as *risky challengers*, because they preferred to win
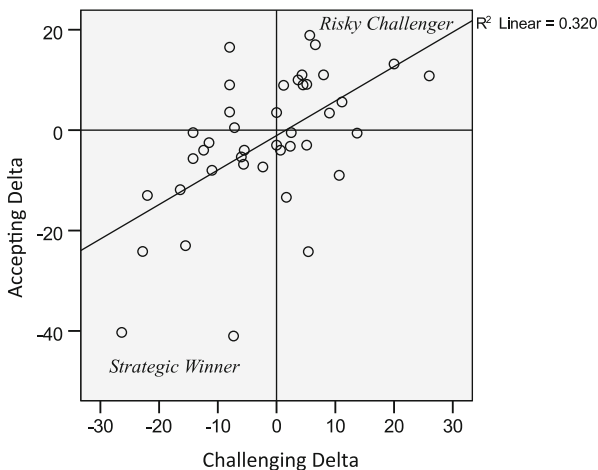
Fig. 5 A scatter plot showing students' average rating difference when they challenged (X-axis) and when they accepted a challenge request from others (Y-axis). On both axes, the difference is computed by subtracting the student's rating from the opponent's rating. The regression coefficient; $b=0.70$

against strong opponents (arguably) to make a big rating leap on a win. On the other hand, the bottom left quadrant shows students who challenged lower rated opponents and accepted challenge requests from lower rated students. This group of students could be labeled as *strategic winners*, because they were (arguably) more focused on winning the game for a small but steady rating accumulation. There were 12 (29.3 %) risky challengers and 19 (46.3 %) strategic winners.

As predicted, there was a significant difference in the final rating between risky challengers and strategic winners; on average, strategic winners attained 18.8 higher final rating than risky challengers. The data also confirmed a clear pattern of *predatory group dynamics*. About 70 % of time, the competition was between a strategic winner and a risky challenger. In other words, strategic winners won the game at the cost of risky challengers' loss.

We hypothesized that risky challengers (RC) would have showed more tutor learning than strategic winners (SW), because risky challengers needed to tutor their SimStudents better than strategic winners to win the game against the higher rated opponents. This hypothesis was not supported. There was no statistically significant difference in tutor learning between the two groups. The average post-test score of procedural skill test; M_{RC}=.43 (SD=0.17) vs. M_{SW}=.52 (SD=0.21). When the post-test score of the procedural skill test was predicted with challenge strategy, strategic winner vs. risky challenger, and the pre-test score of the procedural skill test as a covariate, only the pre-test score was significantly contributed to the regression model; $F(1, 40)$=30.34, $p<0.001$ for pre-test score and $F(1,40)$=0.56, $p=0.58$ for challenge strategy. One explanation is that risky challengers were basing their strategy to increase their rating on the *hope* that their SimStudents would win against stronger SimStudents rather than on *more thorough training* of their SimStudents.

Was there any difference in tutoring engagement between risky challengers and strategic winners? The data did not reveal any notable differences in tutoring engagement between risky challengers and strategic winners in terms of the probability of returning to tutoring after a competition; M_{RC}=.16 (SD=.23) vs. M_{SW}=.15 (SD=.16); $t(25)$=−.223, p=.825. There was no notable difference in the probability of providing "deep" explanations, either. These observations indicate that the two groups of students were equally engaged in tutoring their SimStudents.

Cautious readers might ask how strategic winners became "strategic." Did they involuntarily become strategic winners, because they attained higher rating quicker than other students, which necessarily forced them to compete against lower rated opponents? Or, did strategic winners actually strategically select lower opponents among the candidates with various ratings? To answer this question, we analyzed the preference on selecting opponents as well as the correlation between prior knowledge and challenge strategy.

We first computed students' desire in selecting opponents. We hypothesized that, in the Game Show, students' *desire of challenge* affected their decisions when they requested a challenge to a selected opponent, accepted a challenge request issued by an opponent, or declined a challenge request. If strategic winners strategically selected lower rated opponents, then they should have had a "worse" desire of challenge (as defined below) than risky challengers.

A desire of challenge was coded for each challenge request as either "better" or "worse" in the following manner. A desire of challenge was defined to be "better" when students requested a challenge against a higher rated opponent, accepted a challenge from

a higher rated opponent, and rejected a challenge from a lower rated opponent. A desire of challenge was defined to be "worse" when students requested a challenge against a lower rated opponent, accepted a challenge from a lower rated opponent, and rejected a challenge from a higher rated opponent. For each student, each decision on challenge request was coded and then the degree of worse desire of challenge was computed as a ratio of worse desire decisions to the total number of decisions made. The result showed that strategic winners (SW) had a higher degree of worse desire than risky challengers (RC); M_{SW}=0.66 (SD=0.14) vs. M_{RC}=0.40 (SD=0.12); $F(1, 40)$=9.54, $p<0.001$. This implies that strategic winners had a chance to compete with higher rated opponents but purposefully avoided them.

There was always a top-rated student who technically could not have a "better" desire of challenge. However, the top rated student at any one time was only 2.3 % (1 out of 44) of the entire population and there were many students in or tied for the top position throughout the study. Therefore, the impact of the restriction for the top rated students that they could not actually select higher rated opponents is practically very small in our analysis. (A similar argument applies for students sometimes in the lowest rated position.)

Second, we examined students' final ratings—if strategic winners unintentionally became "strategic," because there were not enough higher rated competitors, then the average final rating should be relatively high with small standard deviations. This hypothesis was not supported. The mean final rating for strategic winners was 32.9 with a standard deviation of 15.0 ranging from 1 to 58. This indicates that there were actually higher rated strategic winners who could have been challenged by strategic winners even towards the end of the study session.

Third, we examined the impact of prior knowledge, measured as the procedural skill pre-test score, and challenge strategy. If a group of students necessarily (hence involuntarily) became strategic winners merely because they attained higher rating quicker than other students, then there should be notable correlation between prior knowledge (pre-test score on the procedural skill test) and challenge strategy. It turned out that there was no correlation between pre-test scores and challenge strategy. There was also no correlation between tutoring accuracy, measured as the percent correct feedback and hint among all feedback and hint provided, and challenge strategy.

In sum, our data suggested that strategic winners selected lower rated opponents purposefully for an easy win. The current design of Game Show therefore contains an unfavorable characteristic for tutor learning—the matching schema for competition for students to select their opponents must be reconsidered.

Discussion

The Effect of Tutor Learning with a Synthetic Peer

Our first hypothesis, that the APLUS environment enhances tutor learning, was partially supported. The study showed that students' proficiency in solving linear equations (measured by the procedural skill test) improved by teaching SimStudent. On the other hand, students' proficiency in conceptual understanding (measured by the conceptual knowledge test) did not improve.

Returning to the question of an underlying cognitive theory of tutor learning, we provide a tentative explanation for how our online learning environment enhances learning by teaching. We saw that the Game Show increased deep responses to SimStudent prompts. These are like self-explanations (Chi et al. 1989) and one might reasonably think they should enhance conceptual (or declarative) learning as in the theory expressed by Aleven and Koedinger (2002). However, we only observed improvements in procedural learning. Further research is needed, but we hypothesize that this effect may be a consequence of improved learning of conditions of applicability (the if-part of a production) of algebraic operations and that these were enhanced by the particular kinds of prompts that SimStudent gave to students. These particular prompts were given to the student when SimStudent made an over-generalization error—applying an operation where it was not appropriate—and SimStudent asks the student why she should not apply that operation in the current context (e.g., "*But before I did 7 for the result of subtract 4 and 7+ 4. I thought that would be the same with 8 here. Why is this different?*"). For example, a Game Show student replied "*Because when you combine like terms 2-3b-8, the like terms 2 and 8 are combined.*" These self-explanations were designed to help students to learn better conditions on their knowledge application. It appears this learning by teaching move (and this particular form of self-explanation) led to student improvement on procedural outcomes as students' procedural knowledge was enhanced in a relatively deep way, through enhanced conditional knowledge. Why such learning did not show up on our conceptual assessment is unclear, but it may suggest the need for conceptual items that better facilitate learning conditional knowledge.

One potential interpretation for the lack of impact on conceptual knowledge is the limited ability of the system to understand students' free-style input. Even when students gave "deeper" responses to SimStudent's questions, SimStudent was not actually benefitting more from these responses, because it could not make use of them. Without feedback, the student tutors did not make the conceptual gains that may be possible with a more interactive dialogue. It may be that students can learn from attempts to self-explain even without feedback (cf., Chi et al. 1989), but we did not observe such an effect here on conceptual outcomes.

Motivation and Engagement in the Context of Competitive Game Show

The second hypothesis about the tutoring engagement was supported. The data confirmed that Game Show students showed greater engagement in tutoring SimStudent as indicated by a higher probability of responding to SimStudent questions more often and deeply. Although both conditions showed the same motivation level measured as the self-reported response to the post-study questionnaire, students in the Game Show condition showed greater engagement in tutoring their SimStudents than the Baseline students (as shown in the process data).

The observations about shorter time on task for the Game Show students were expected, because Game Show students spent their time on the Game Show when the total amount of study duration was the same as the baseline condition. However, the fact that *the Game Show students achieved the same test score as the Baseline students by tutoring fewer problems (while spending the same tutoring time per problem) is notable.*

As for the desire to win, there was a clear predatory pattern between risky challengers and strategic winners such that about 70 % of strategic winners' wins

were against risky challengers. This indicates that some students had stronger desire to win by easy competition than others, and their performance goals did not line up to learning goals.

Impact of the Motivation and Engagement on Tutor Learning

Our third hypothesis on the impact of the tutoring engagement on tutor learning was not supported. There was no condition difference in test scores both for procedural skill test and conceptual knowledge test. The data did not confirm any notable correlation between tutoring engagement and tutor learning. At the same time, our data did not confirm any negative influence of the desire to win, i.e., the external motivation, against tutor learning. There was no notable difference in tutor learning between risky challengers and strategic winners.

A striking fact is that there was no difference in tutoring engagement between risky challengers and strategic winners (although the data showed that the Game Show students were more engaged in tutoring than Baseline students as mentioned above). Risky challengers in our study, who should have been motivated to make their SimStudent a stronger competitor, did not actually tutor any better than strategic winners (or, conversely, strategic winners were equal-level tutors with risky challengers). This may be partly due to the lack of tutoring after Game Show.

Design Implications for Future Studies

Our data suggest that there is no overall negative evidence that the proposed competitive Game Show is detrimental for tutor learning. Rather, the data imply that the Game Show has both positive (as students would be more engaged in tutoring) and negative (as students would have shorter time on tutoring) aspects for tutor learning.

The data showed that being a strategic winner (that is, having a high extrinsic, competitively-oriented motivation) is not necessarily undesirable in our learning environment. The issue is rather at the misalignment between the performance goal and learning goal as well as the lack of alternating the competition and tutoring especially after a loss of competition.

One idea to align the goals of the Game Show and learning is to ensure that the winners' SimStudents have a high competency of solving the target problems. One such idea is to restrict the range of available opponents so that the difference in ratings is within a desired zone. In this regard, our data actually support Miller et al.'s (1999) claim that such a *knowledge-dependency*, i.e., the relation between the pedagogically targeted concepts and the knowledge required to interact successfully with the game environment, is key for successful learning. Likewise, Biggs and Tang (2007) pointed out the importance of the *constructive alignment*, which basically claims that students learn what they do. In the game-show used in the Betty's Brain system, the game goal was directly correlated with the learning goal (Schwartz et al. 2009).

To resolve the lack of tutoring, Game Show should provide equal learning opportunities for all students regardless of the result of the competition. One simple idea is to make it a requirement that students need to tutor their SimStudent after a loss. Embedding virtual Game Show contestants with various levels of competency and setting the goal of Game Show to beat the virtual contestants would encourage

students (especially strategic winners) to tutor their SimStudents better. Since we can pre-train SimStudent to different extents and types of problems, making such virtual contestants is technically feasible.

Conclusion

We found that introducing the competitive Game Show to the learning environment increased students' engagement in tutoring and desire to win the competition, but in order to connect engagement and motivation to tutor learning, the goals of the Game Show must be better aligned with the learning goal.

As a major technological contribution of the current study, we have demonstrated that SimStudent can be used as a teachable agent with which students can learn procedural skills by teaching. We also demonstrated that using a teachable agent allows researchers to collect detailed process data that, when combined with outcome data, can help us understand better cognitive theories of learning by teaching. We also identified issues of the current Game Show design and proposed design improvements based on the process data and learning outcome data.

The SimStudent technology provides a promising platform for experimenting with approaches like teachable agents and learning companions. Such studies can be run in real schools and these approaches have the potential to provide effective interventions. The transformative theory of learning by teaching has yet to be thoroughly investigated, but we find great potential and opportunity in future work in this area.

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*(2), 147–179.

Bernacki, M. L., Nokes-Malach, T. J., & Aleven, V. (2013). Fine-grained assessment of motivation over long periods of learning with an intelligent tutoring system: Methodology, advantages, and preliminary results. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 629–644). New York: Springer.

Biggs, J., & Tang, C. (2007). *Teaching for quality learning at university* (3rd ed.). New York: McGraw-Hill.

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

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: Cognitive Science Society.

Bredeweg, B., Gómez-Pérez, A., André, E., & Salles, P. (2009). DynaLearn - engaging and informed tools for learning conceptual system knowledge. In P. R. R. Azevedo & G. Biswas (Eds.), *Cognitive and Metacognitive Educational Systems (MCES2009), AAAI Fall Symposium* (pp. 46–51). Arlington: AAAI Press.

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.

Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inference and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology: Educational design and cognitive science* (Vol. 5, pp. 161–238). Mahwah: Lawrence Erlbaum Associates.

Chi, M. T., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: how students study and use examples in learning to solve problems. *Cognitive Science, 13*(2), 145–182.

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

Cohen, E. G. (1994). Restructuring the classroom: conditions for productive small groups. *Review of Educational Research, 64*(1), 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*(2), 237–248.

Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: modeling the acquisition of procedural knowledge. *User Modeling and User Adapted Interaction, 4*(4), 253–278.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology, 18*(1), 105–115. doi:10.1037/h0030644.

Deci, E. L., Betley, G., Kahle, J., Abrams, L., & Porac, J. (1981). When trying to win: competition and intrinsic motivation. *Personality and Social Psychology Bulletin, 7*(1), 79–83. doi:10.1177/014616728171012.

Devin-Sheehan, L., Feldman, R. S., & Allen, V. L. (1976). Research on children tutoring children: a critical review. *Review of Educational Research, 46*(3), 355–385. doi:10.2307/1170008.

Elo, A. E. (1978). *The rating of chessplayers, past and present*. New York: Arco.

Gartner, A., Kohler, M., & Riessman, F. (1971). *Children teach children: Learning by teaching*. New York: 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*(6), 495–522.

Hedin, D. (1987). Students as teachers: a tool for improving school. *Social Policy, 17*(3), 42–47.

Koedinger, K. R., Aleven, V., Roll, I., & Baker, R. (2009). In vivo experiments on whether supporting metacognition in intelligent tutoring systems yields robust learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 897–964). New York: Routledge.

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

Matsuda, N., Cohen, W. W., Sewall, J., Lacerda, G., & Koedinger, K. R. (2007). Predicting students performance with simstudent that learns cognitive skills from observation. In R. Luckin, K. R. Koedinger, & J. Greer (Eds.), *Artificial intelligence in education* (pp. 467–476). Amsterdam: Ios Press.

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: 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: Cognitive Science Society.

Matsuda, N., Keiser, V., Raizada, R., Tu, A., Stylianides, G. J., Cohen, W. W., & Koedinge, K. R. (2010). Learning by teaching simstudent: Technical accomplishments and an initial use with students. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Proceedings of the international conference on intelligent tutoring systems* (pp. 317–326). Heidelberg: Springer.

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: Springer.

Matsuda, N., Cohen, W. W., Koedinger, K. R., Keiser, V., Raizada, R., Yarzebinski, E., & Stylianides, G. J. (2012a). 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: IEEE Computer Society.

Matsuda, N., Yarzebinski, E., Keiser, V., Raizada, R., Stylianides, G. J., & Koedinger, K. R. (2012b). 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: Springer.

Matsuda, N., Yarzebinski, E., Keiser, V., Raizada, R., William, W. C., Stylianides, G. J., & Koedinge, K. R. (2013). Cognitive anatomy of tutor learning: Lessons learned with SimStudent. *Journal of Educational Psychology.* doi:10.1037/a0031955.

Miller, C. S., Lehman, J. F., & Koedinger, K. R. (1999). Goals and learning in microworlds. *Cognitive Science, 23*(3), 305–336.

Mitchell, T. M. (1997). *Machine learning*. Boston: McGraw-Hill.

Muggleton, S., & de Raedt, L. (1994). Inductive logic programming: theory and methods. *Journal of Logic Programming, 19-20*(Supplement 1), 629–679.

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: Springer.

Quinlan, J. R. (1990). Learning logical definitions from relations. *Machine Learning, 5*(3), 239–266.

Reeve, J., & Deci, E. L. (1996). Elements of the competitive situation that affect intrinsic motivation. *Personality and Social Psychology Bulletin, 22*(1), 24–33. doi:10.1177/0146167296221003.

Reeve, J., Olson, B. C., & Cole, S. G. (1985). Motivation and performance: two consequences of winning and losing in competition. *Motivation and Emotion, 9*(3), 291–298. doi:10.1007/BF00991833.

Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive tutor: applied research in mathematics education. *Psychonomic Bulletin & Review, 14*(2), 249–255.

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*(4), 534–574.

Russell, S. J., & Norvig, P. (2003). *Artificial intelligence: A modern approach* (2nd ed.). Upper Saddle River: Prentice Hall.

Schwartz, D. L., Chase, C., Chin, D. B., Oppezzo, M., Kwong, H., Okita, S. Y., & Wagster, J. D. (2009). Interactive metacognition: Monitoring and regulating a teachable agent. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 340–358). New York: Routledge.

VanLehn, K., Ohlsson, S., & Nason, R. (1994). Applications of simulated students: an exploration. *Journal of Artificial Intelligence in Education, 5*(2), 135–175.

Vansteenkiste, M., & Deci, E. L. (2003). Competitively contingent rewards and intrinsic motivation: can losers remain motivated? *Motivation and Emotion, 27*(4), 273–299. doi:10.1023/a:1026259005264.