A Comparison of Gains between Educational Games and a Traditional ITS

G. Tanner Jackson[1], Chutima Boonthum-Denecke[2], and Danielle S. McNamara[1]
[1]Learning Sciences Institute, Arizona State University
TannerJackson@asu.edu, Danielle.McNamara@asu.edu
[2]Department of Computer Science, Hampton University
chutima.boonthum@hamptonu.edu

Abstract

Intelligent Tutoring Systems (ITSs) have begun to incorporate game-based components in an attempt to balance the learning benefits of ITSs with the motivational benefits of games. iSTART-ME (Motivationally Enhanced) is a new game-based learning environment that was developed on top of an existing ITS (iSTART). In a multi-session lab-based efficacy study with 125 high school students, those students with a low prior reading ability who were trained by a game-based tutoring system (iSTART-ME) or a traditional intelligent tutoring system (iSTART-Regular) performed significantly better on posttest measures than students assigned to a time-delayed control condition. Additionally, the low reading ability students who interacted with the game-based system had a tendency to gain more than students in the traditional ITS system.

Introduction[1]

Intelligent Tutoring Systems (ITSs) have been producing consistent learning gains for decades. However, a common problem with these systems is maintaining student engagement without reducing the learning benefits. This is particularly problematic for ITSs that require long-term tutorial interactions that span across days, weeks, or even months. Student interest within these types of tutoring systems often wanes over time due to the repetitive nature of practice tasks.

One previously successful solution to improve engagement has been to incorporate game-like components into educational environments (for a review, see Clark, Nelson, Sengupta, & D'Angelo, 2009). Several systems have taken this route and begun to create combinations of Intelligent Tutoring and Games (ITaG; McNamara, Jackson, & Graesser, 2010). One example of this endeavor is the Interactive Strategy Training for Active Reading and Thinking-Motivationally Enhanced (iSTART-ME) which was built on top of an existing ITS (called iSTART) and adapted into a game-based environment where students can practice strategies, earn points, advance through levels, purchase rewards, create a personalized avatar, and play educational mini-games. The remainder of this work describes the two iSTART systems and discusses an experimental comparison of the two environments.

iSTART

iSTART is an ITS designed to improve students' reading comprehension by teaching self-explanation in combination with effective reading strategies. iSTART introduces students to the concept of self-explanation and provides instruction on how to use reading comprehension strategies to improve their understanding of difficult science texts. The development of iSTART was based on previous research with a successful human intervention called SERT (Self-Explanation Reading Training: McNamara, 2004; O'Reilly, Taylor, & McNamara, 2006). This training was designed to help those low ability students who might not effectively use strategies on their own. Students who have been provided with iSTART have shown significant improvement in reading comprehension, comparable to the performance within SERT (Magliano, Todaro, Millis, Wiemer-Hastings, Kim, & McNamara, 2005). iSTART training is separated into three distinct modules that instantiate the pedagogical principle of modeling-scaffolding-fading: introduction, demonstration, and practice, respectively.

During the introduction module, three animated agents (one teacher and two students) hold a vicarious classroom-like dialogue. This dialogue presents the concept of self-explanation and the associated iSTART reading strategies (comprehension monitoring, prediction, paraphrasing, elaboration, and bridging). These agents interact with one another to provide descriptions, examples, and counter examples of each reading strategy. After each strategy discussion, formative assessments are presented that gauge the student's current level of understanding for that strategy.

After all strategies have been introduced and modeled, the system transitions into the demonstration module. The demonstration module utilizes two animated agents (one teacher, one student) that apply the self-explanation strategies to an example text. During this scaffolding phase the user is asked to analyze and identify the various strategies being used by the student agent. The dialogue and feedback between the animated agents foreshadows the interaction that the users will have during the practice module.

The practice module in iSTART affords students the opportunity to apply the self-explanation strategies within their own self-explanations. This module fades out most direct instruction and uses formative feedback to guide the interaction. Merlin (the teacher agent during demonstration) serves as the self-explanation coach by providing feedback for every student-generated self-explanation and prompting them to use the newly acquired strategies. The main purpose of this module is to provide students with an opportunity to apply the strategies to new texts and to integrate knowledge from different sources in order to understand a challenging text.

During practice, each self-explanation that a student generates is assessed by the iSTART assessment algorithm. This assessment helps to inform the feedback provided by the system. The algorithm output is coded as a 0, 1, 2, or 3. An assessment of "0" relates to self-explanations that are either too short or contain mostly irrelevant information. An iSTART score of "1" is associated with a self-explanation that primarily relates only to the target sentence itself (sentence-based). A "2" means that the student's self-explanation incorporated some aspect of the text beyond the target sentence (text-based). If a self-explanation earns a "3", then it is interpreted to have incorporated information at a global level, and may include outside information or refer to an overall theme across the whole text (global-based). This algorithm has demonstrated performance comparable to that of humans, and provides a general indication of the cognitive processing required to generate a self-explanation (Jackson, Guess, & McNamara, 2010).

Within iSTART there are two types of practice modules. The first practice module is situated within the core context of iSTART (initial 2-hour training) and includes two texts. The second practice module is a form of extended interaction, and it operates in the exact same manner as the original practice module. During this extended practice phase, a teacher can assign specific texts for students to read. These texts are either already in the system or can be added to the system on short notice. Because of the need to incorporate various texts, the iSTART feedback algorithm has been designed to adapt to new texts, and its performance has been comparable to that of humans (Jackson, Guess, McNamara, 2010). The extended practice module is designed to provide a long-term learning environment that can span weeks or months. Research on iSTART has shown that the extended practice effectively increases students' performance over time (Jackson, Boonthum, & McNamara, 2010). However, one unfortunate side effect of this long-term interaction is that students often become disengaged and uninterested in using the system (Bell & McNamara, 2007).

iSTART-ME

To combat the problem of disengagement over time, the iSTART extended practice module has been situated within a game-based environment called iSTART-ME (motivationally enhanced). This game-based environment builds upon the existing iSTART system and was specifically designed to increase persistence and active engagement for low ability students who are more likely to disengage from extended training. The iSTART-ME system and design rationale has been more extensively described in other papers, so only the relevant aspects will be described here (Jackson, Boonthum, & McNamara, 2009; Jackson, Dempsey, & McNamara, 2010).

The main focus of the iSTART-ME project is to implement and assess game-based principles and features that are expected to support effective learning, increase motivation, and sustain engagement throughout a long-term interaction within an established ITS. Previous research has indicated that increasing self-efficacy, interest, engagement, and self-regulation should positively impact learning (Alexander, Murphy, Woods, Duhon, & Parker, 1997; Bandura, 2000; Pajares, 1996; Pintrich, 2000; Zimmerman & Schunk, 2001). The iSTART-ME project attempts to manipulate these motivational constructs via game-based features that map onto one of the following five categories: feedback, incentives, task difficulty, control, and environment. These categories are discussed in detail in (McNamara, Jackson, & Graesser, 2010).

The previous version of iSTART automatically progressed students from one text to another with no intervening actions. The new version of iSTART-ME is controlled through a selection menu (see Figure 1 for screenshot). This selection menu provides students opportunities to interact with new texts, earn points, advance through levels, purchase rewards, personalize a character, and play educational mini-games (designed to use the same strategies as in practice).

Figure 1. Screenshot of iSTART-ME selection menu.

Within iSTART-ME, students can earn points as they interact with texts and provide their own self-explanations (top of Figure 1). Each time that a student submits a self-explanation, it is assessed by the iSTART algorithm and points are awarded based on a scoring rubric. The rubric has been designed to reward consistent good performance. So, students earn more points if they repeatedly provide good self-explanations on consecutive turns, but earn fewer points if they fluctuate between good and poor performance. In addition to providing a form of feedback, earning points within iSTART-ME serves three main purposes: advancing through levels, purchasing rewards, and unlocking menu features.

As students accumulate more points, they advance through a series of levels, and each new level unlocks one or more new features or games. Each subsequent level requires an increasing number of points, therefore students must expend slightly more effort to achieve further advancements. The levels are labeled to help increase interest (e.g., "ultimate bookworm", "serious strategizer", etc.), and also help to serve as global indicators of progress across texts.

Points can also be used to "purchase" rewards within the system (bottom box in Figure 1). One of the options available as a reward is for students to change aspects of the learning environment. They can spend some of their iBucks to choose a new tutor agent, change the interface to a new color scheme, or update the appearance of their personal avatar. These features provide students with a substantial amount of control and personalization, and have been designed as purchasable replacements, rather than always available options, to help reduce off-task behaviors (such as switching back and forth between agents).

Lastly, a suite of eight educational mini-games have been designed and incorporated within the iSTART-ME extended practice module. Some mini-games require identification of the type of strategy use, while others may require students to generate their own self-explanations. The majority of iSTART-ME mini-games require similar cognitive processes enveloped within different combinations of gaming elements.

Figure 2. Screenshot of Showdown.

Showdown and Map Conquest are two methods of generative game-based practice that use the same iSTART assessment algorithm from regular practice. In Showdown (see Figure 2), students compete against a computer player to win rounds by writing better self-explanations. After the learner submits a self-explanation, it is scored, the quality assessment is represented as a number of stars (0-3), and an opponent self-explanation is also presented and scored. The self-explanation scores are compared and the player with the most stars wins the round. The player with the most rounds at the end of the game is declared the winner. Map Conquest is the other game-based method of practice where students generate their own self-explanations. In this game the quality of a student's self-explanation determines the number of dice that student earns. Students place these dice on a map, and use them to conquer neighboring opponent territories, which are controlled by two virtual opponents.

In most of the identification mini-games, for example Balloon Bust (Figure 3), students are presented with a target sentence and an example self-explanation. The student must decide which iSTART strategy was used in the self-explanation and then click on the corresponding balloons. There are also three other mini-games that focus on the same task of identifying strategies within example self-explanations. These other games each incorporate a new interface with a different combination of game elements, including fantasy, competition, and perceptual aspects (as in Balloon Bust). Though the surface features of these games can differ widely, they have been designed with very similar leveling structures and can all be completed within 10-20 minutes. Students are allowed to select any form of practice or mini-game from the selection menu that has been unlocked (provided that they have enough iBucks).

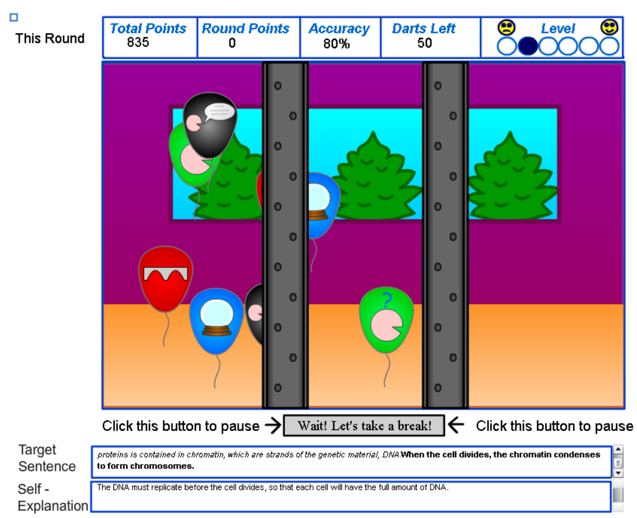

Figure 3. Screenshot of Balloon Bust.

Current Study

Previous research focusing on components of the iSTART-ME system yielded somewhat conflicting patterns of results depending on the time-scale of the intervention. This research indicated that after a short-term interaction (~60 minutes, including brief training), students who used a game-based method of practice performed worse than students using a non-game-based environment (Jackson, Dempsey, & McNamara, in press). However, in a longer-term pilot evaluation with full training (~6 hours across multiple sessions), students performed equally well using either the game-based or the non-game-based practice environments. Therefore, one possible concern with integrating games into learning systems is that they have the potential to detract from the immediate pedagogical goals and reduce learning improvements in the short-term. However, across long-term training, the engagement fostered by the game environment may compensate for any distracting elements, thus allowing students to catch up in performance (Jackson, Dempsey, & McNamara, in press).

The current work was conducted to expand upon the findings from the previous research and to more thoroughly explore the potential long-term benefits of game-based training. This study includes a multi-session experiment which directly compares the effectiveness of the newly designed educational game system (iSTART-ME) to a traditional intelligent tutoring system (iSTART-Regular) and to a control group that received no tutoring (control).

Procedure

One hundred and twenty five high school students from a mid-south urban environment participated in a 10-session experiment involving three phases: pretest, training, and posttest. During the first session, students completed a pretest that included measures to assess their attitudes (towards reading, technology, science, games), prior knowledge (science domain and general vocabulary), and prior abilities (reading comprehension and self-explanation). During the training phase participants were randomly assigned into one of three conditions: iSTART-ME, iSTART-Regular, and control. Students within the iSTART-ME condition interacted with the full game-based selection menu across 8 separate sessions of at least 1 hour each. Participants in the iSTART-Regular condition used the original non-game-based version of iSTART for the same amount of time (8 sessions of at least 1 hour each). Students in the control condition received no training until after they completed a one week delayed posttest. All students completed the posttest, which included measures similar to the pretest.

The training within both iSTART conditions was identical until the participants transitioned into extended practice. The iSTART-ME students then used the full selection menu, while the iSTART-Regular students continually transitioned from one text to another all within the regular practice environment. Both systems allowed students to progress through the tutoring at their own pace and therefore not all students experienced the same components at the same time. Several students completed the regular practice and transitioned into the extended practice during the second session, while other students may not have reached the extended practice section until the third or possibly fourth sessions. Ultimately, all students completed the training modules and subsequently interacted with their randomly assigned training condition for the remainder of the study (totaling 8 full training sessions of at least 1 hour each).

Results

The pretest and posttest outcomes were analyzed to assess the potential impacts of the three training conditions and students' prior reading abilities. The following analyses focus on the differences in students' self-explanation quality from pretest to posttest.

A median split using the pretest Gates-MacGinitie reading comprehension test was used to create groups of students with either high or low prior reading ability. Those students in the high reading ability group had a significantly higher proportion of correct answers on the comprehension test (M=.64) than did the students in the low reading ability group (M=.30), $F(1,123)=265.10$, $p<.01$.

An ANOVA conducted on the pretest self-explanation scores indicated that the participants with lower prior reading ability generated significantly lower quality self-explanations than the high ability students, $F(1,119)=16.76$, $p<.01$ (see Figure 4 for pretest means). There were no significant differences of pretest self-explanation quality between the three conditions, $F(2,119)=0.75$, $p=.47$, nor was there a significant interaction of pretest self-explanation quality between condition and prior reading ability, $F(2,119)=0.52$, $p=.60$.

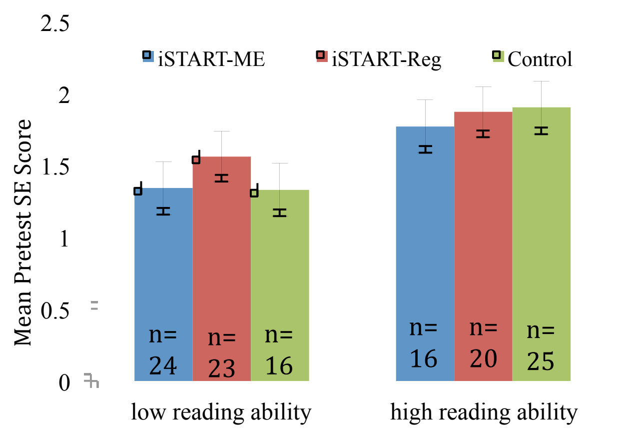

Figure 4. Mean Pretest Self-Explanation (SE) Scores.

Analyses conducted on the posttest self-explanation quality found significant differences between conditions, but no differences between reading ability or an interaction (see Figure 5 for posttest means). Specifically, an ANOVA on posttest self-explanation quality revealed that both training conditions (iSTART-Regular and iSTART-ME) produced significantly higher quality posttest self-explanations than the control condition, $F(2,119)=10.07$,

$p<.01$. There was a marginally significant difference of posttest self-explanation quality where high ability students produced slightly better quality posttest self-explanations than the participants with low prior reading ability, $F(1,119)=3.41$, $p=.07$. There was not a significant interaction of quality between condition and reading ability, $F(2,119)=0.78$, $p=.46$.

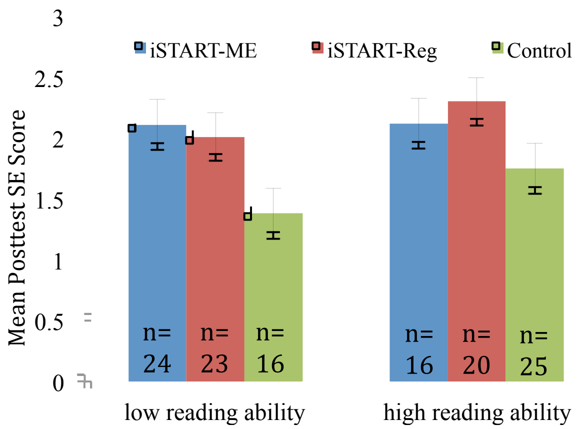

Figure 5. Mean Posttest Self-Explanation (SE) Scores.

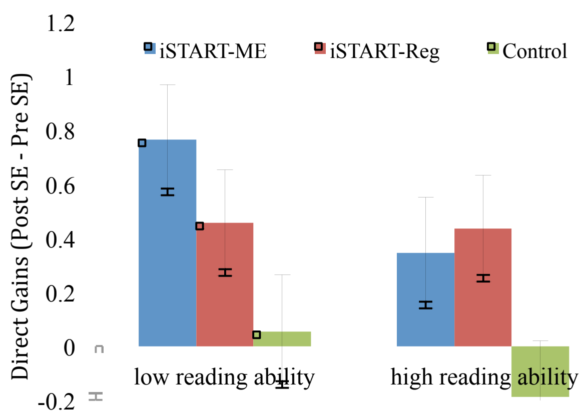

Figure 6. Direct Gains on Self-Explanation (SE) Scores.

A direct gains score was calculated to further investigate learning differences between conditions. A gains score was calculated for each participant by subtracting their pretest self-explanation score from their posttest self-explanation score (see Figure 6 for mean gains scores). An ANOVA on the direct gains scores yielded a significant main effect for condition, $F(2,119)=9.66$, $p<.01$, a marginal effect for reading ability, $F(1,119)=3.53$, $p=.06$, and a non-significant interaction, $F(2,119)=0.96$, $p=.39$. Both versions of the iSTART training produced significantly higher gains than the control condition. The participants with lower prior reading ability gained slightly more than the high reading ability students, though they started with lower pretest scores and therefore had more room for improvement. The interaction between condition and reading ability was not significant. However, the iSTART-ME group had a tendency to produce the most gains within the target ability group (i.e., students with low prior ability).

Conclusions

The overarching goal of the iSTART-ME project has been to further our understanding of the benefits of adding game-based elements to ITSs (coined ITaG, Intelligent Tutoring and Games in McNamara et al., 2009). The current efficacy study was the first major step in assessing a fully implemented ITaG built on top of an existing ITS. This study was designed to compare the potential learning benefits between a traditional ITS (iSTART-Regular), a newly developed ITaG system (iSTART-ME), and a time-based control condition.

Results indicate that the students' prior reading ability is related to their initial self-explanation performance at pretest (higher reading ability yields higher quality pretest self-explanations). However, this relation decreased after training (became only marginally significant) and the quality of self-explanations at posttest was instead related to the randomly assigned training condition, with both iSTART systems outperforming the control condition regardless of initial reading ability. The analyses on direct gains for self-explanation improvement indicated that all students who received training improved above and beyond the control group, as expected. Though the interaction was not significant, the largest improvements in self-explanation quality were found for the low ability readers within the iSTART-ME condition (and these low ability readers are the targeted end-users of the training system). These results support the main goal of the project, and provide further evidence that games can be effectively integrated within a learning environment.

The current long-term evaluation goes well beyond immediate short-term findings to explore the effects of games during prolonged skill acquisition (i.e., using comprehension strategies effectively within self-explanations). During this extended interaction, students within the iSTART-ME condition had access to the full selection menu and therefore could spend more time off-task interacting with various features and mini-games. In contrast, students within the iSTART-Regular condition had continued generative practice and received formative feedback on how to generate higher quality self-explanations. Despite the potential differences in time-on-task allocation, the students using the game-based system gained equivalently to (and even slightly more than) students in the traditional ITS. This finding supports the long-term learning trend from our previous work (Jackson, Dempsey, & McNamara, in press), and creates a promising foundation from which we can extend subsequent work and further contribute to the scientific research on game-based learning.

The development of iSTART-ME allows us to examine the effectiveness of a combined ITaG system, as well as to more systematically evaluate the effects of game components in the context of an ITS. The current system has been designed with distinct and separable features so that multiple combinations can be tested across a variety of experiments. Future work with iSTART-ME includes both global and local assessments of game-based performance. Further analyses of this efficacy study are anticipated that will investigate user performance, enjoyment, attitudes, engagement, and persistence across time. Additionally, several small-scale experiments are being implemented to address the interactions between specific game components.

Both the current and future work of iSTART-ME helps to further the field of Intelligent Tutoring Systems and game-based learning. The design of iSTART-ME is very modular and provides an interesting theoretical alternative to the growing number of fully immersive epistemic games. Ultimately, we expect hybrid ITaG learning environments to dramatically impact the effectiveness of computer-based training as well as further our understanding of the complex motivational aspects of learning environments and their interplay with learning.

Acknowledgments

This research was supported in part by the Institute for Educational Sciences (IES R305G020018-02; R305G040046, R305A080589) and National Science Foundation (NSF REC0241144; IIS-0735682). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the IES or NSF.

References

Alexander, P.A., Murphy, P.K., Woods, B.S., Duhon, K.E., and Parker, D. 1997. College instruction and concomitant changes in students' knowledge, interest, and strategy use: A study of domain learning, *Contemporary Educational Psychology* 22: 125-146.

Bandura, A. 2000. Self-efficacy: The foundation of agency. In W. Perig and A. Grob (Eds.). *Control of human behavior, mental processes, and consciousness: Essays in honor of the 60th birthday of August Flammer,* Mahwah, NJ: Erlbaum, 17-33.

Bell, C., and McNamara, D.S. 2007. Integrating iSTART into a high school curriculum. In *Proceedings of the 29th Annual Meeting of the Cognitive Science Society, 809-814*. Austin, TX: Cognitive Science Society.

Clark, D., Nelson, B., Sengupta, P., D'Angelo, C. 2009. *Rethinking science learning through digital games and simulations: Genres, examples, and evidence*. Washington DC: National research Council.

Jackson, G.T., Boonthum, C., and McNamara, D.S. 2009. iSTART-ME: Situating extended learning within a game-based environment. In H.C. Lane, A. Ogan, and V. Shute (Eds.), In *Proceedings of the Workshop on Intelligent Educational Games at the 14th Annual Conference on Artificial Intelligence in Education,* 59-68. Brighton, UK: AIED.

Jackson, G.T., Boonthum, C., and McNamara, D.S. 2010 The efficacy of iSTART extended practice: Low ability students catch up. In J. Kay and V. Aleven (Eds.), In *Proceedings of the 10th International Conference on Intelligent Tutoring Systems*, 349-351. Berling/Heidelberg: Springer.

Jackson, G.T., Dempsey K.B., and McNamara, D.S. 2010. The evolution of an automated reading strategy tutor: From classroom to a game-enhanced automated system. In M.S. Khine and I.M. Saleh (Eds.), *New Science of learning: Cognition, computers and collaboration in education,* 283-306. New York, NY:Springer.

Jackson, G.T., Dempsey, K.B., and McNamara, D.S. in press. Game-based practice in reading strategy tutoring system: Showdown in iSTART-ME. In H. Reinders (Ed.), *Computer games*. Bristol, UK: Multilingual Matters.

Jackson, G.T., Guess, R., and McNamara D.S. 2010. Assessing cognitively complex strategy use in an untrained domain. *Topics in Cognitive Science 2:* 127-137.

Magliano, J.P., Todaro, S., Millis, K., Wiemer-Hastings, K., Kim, H.J., and McNamara, D.S. 2005. Changes in reading strategies as a function of reading training: A comparison of live and computerized training. *Journal of Educational Computing Research 32*: 185-208.

McNamara, D.S. 2004. SERT: Self-explanation reading training. *Discourse Processes 38*: 1-30.

McNamara, D.S., Jackson, G.T., and Graesser, A.C. 2010. Intelligent tutoring and games (ITaG). In Y.K. Baek (Ed.), *Gaming for classroom-based learning: Digital role-playing as a motivator of study*, 44-65. Hershey, PA: IGI Global.

O'Reilly, T., Taylor, R.S., and McNamara, D.S. 2006. Classroom based reading strategy training: Self-explanation vs. reading control. In R. Sun and N. Miyake (Eds.), In *Proceedings of the 28th Annual Conference of the Cognitive Science Society,* 1887. Mahwah, NJ: Erlbaum.

Pajares, F. 1996. Self-efficacy beliefs in academic settings. *Review of Educational Research 66*: 543-578.

Pintrich, P.R., 2000. Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology 92*: 544-555.

Zimmerman, B.J., and Schunk, D.H. 2001. Reflections on theories of self- regulated learning and academic achievement. In B.J. Zimmerman, and D.H. Schunk, (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives,* 289-307. Mahwah, NJ: Erlbaum.