

Self-Regulation and Gender Within a Game-Based Learning Environment

John L. Nietfeld, Lucy R. Shores, and Kristin F. Hoffmann
North Carolina State University

In this study, we examined how self-regulated learning (SRL) and gender influences performance in an educational game for 8th-grade students ($N = 130$). Crystal Island–Outbreak is an immersive, inquiry-based, narrative-centered learning environment featuring a microbiology science mystery aligned with 8th-grade science curriculum. SRL variables predicted successful in-game performance even after accounting for prior knowledge and perceived gaming skill. Content learning gains were found across both genders, and girls performed at similar levels as boys in the game despite incoming disadvantages for perceived skill and prior gaming experience. Boys were more effective than girls in using a cognitive tool that was critical for solving the Crystal Island–Outbreak mystery; however, these differences disappeared when prior gaming experience was taken into account. Overconfidence on monitoring judgments for boys but not girls was predictive of in-game performance. Findings related to motivational variables such as self-efficacy, situational interest, and goal orientation were mixed with regard to their consistency across genders.

Keywords: self-regulated learning, gender, game-based learning, meta-cognition, motivation

Educational computer games are becoming increasingly popular in today's schools (Tobias & Fletcher, 2007). Therefore, it is important that such games encourage self-regulated learning (SRL; Nietfeld & Shores, 2010) and are informed by sound instructional methods (Moreno & Mayer, 2005). Without rigorous testing, such gaming environments will likely be dismissed as “motivational fluff” (O’Neil, Wainess, & Baker, 2005, p. 456) rather than as advances in learning technology. Recently, the evidence for serious games and learning has been promising (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012; Wouters, van Nimwegen, Oostendorp, & van der Spek, 2013). One major advantage of educational gaming environments is their potential to provide a customized experience with opportunities for real-time feedback (Shute & Zapata-Rivera, 2012). Models of SRL can inform this customization by directing the focus to the learner’s strategy use, metacognitive skills, and motivation. However, a focus on models of effective learning, coupled with the need for customization, requires a parallel focus on individual difference variables that might impact learning outcomes. Gender is one such variable that has garnered significant attention within computer-based learning applications (Carr & Pelletier, 2009; Hayes, 2011; Plant, Baylor, Doerr, & Rosenberg-Kima, 2009). Nevertheless, there is still much to be learned about how gender impacts one’s ability to regulate

strategically, metacognitively, and motivationally in complex, inquiry-based environments. The goal of the current investigation was to test the pedagogical efficacy of one game-based learning environment as a platform to examine how SRL and gender influences performance in such environments.

Crystal Island–Outbreak, the game-based learning environment employed in the current study, draws on elements emphasized in computer games such as being rule-based, responsive, challenging, and cumulative (Mayer, 2011) but is unique in that it utilizes a narrative-centered approach in order to encourage and scaffold self-regulation and content learning. Narrative approaches present “story-centric” problem-solving activities where learners are immersed in a narrative tailored to the curriculum (Rowe, Shores, Mott, & Lester, 2010). For example, established multiuser virtual environments such as Quest Atlantis (Barab, Dodge, Thomas, Jackson, & Tuzun, 2007) and River City (Ketelhut, Dede, Clarke, Nelson, & Bowman, 2007) use rich narrative settings to contextualize inquiry-based learning scenarios with strong social and ethical dimensions. Crystal Island–Outbreak takes a similar form as it examines learning within an engaging mystery that requires knowledge of microbiology to solve. Furthermore, by capitalizing on aspects of the narrative, this approach allows for a less invasive means by which to introduce learning scaffolds.

Self-Regulated Learning and Gender in Game-Based Learning Environments

Self-regulated learning refers to the effective regulation of one’s own learning in the pursuit of personal goals. SRL is contextual in nature and changes dynamically in response to “episodes” experienced by the learner (Winne, 2010). We situated our approach within the literature with the belief that SRL is a malleable, multilevel phenomenon informed by models that incorporate aspects of both information processing and social cognitive perspectives (Pintrich, 2000; Winne & Hadwin, 1998; Zimmerman, 2000).

This article was published Online First June 16, 2014.

John L. Nietfeld, Lucy R. Shores, and Kristin F. Hoffmann, Curriculum, Instruction, & Counselor Education, North Carolina State University.

This research was supported by National Science Foundation Grant ALT 0632450. Any opinions, findings, and conclusion or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Correspondence concerning this article should be addressed to John Nietfeld, North Carolina State University, Curriculum, Instruction, & Counselor Education, 602D Poe Hall, Raleigh, NC 27695. E-mail: john_nietfeld@ncsu.edu

At the broadest level, models of SRL are composed of strategic, metacognitive, and motivational components (Zimmerman, 2000).

The ability to self-regulate is positively associated with increased learning, academic performance, and academic motivation (Pintrich, 2000) and is due in large part to one's ability to effectively monitor cognitive strategy use (Butler & Winne, 1995). Autonomy and control are also key aspects of SRL that provide the potential for learners to regulate, monitor, and control their cognition, motivation, and affect (Pintrich, 2000). The degree to which this effective cognitive management is employed is due largely to the motivation to engage these control processes. In open-ended games, such control by the learner is important because students can choose to engage or disengage from academic goals and can solve problems with many different approaches.

To date, there has been a limited amount of research targeting SRL within game-based learning environments. A review by O'Neil, et al. (2005) of 19 peer-reviewed empirical studies examining the effectiveness of digital games found that none categorically addressed SRL. Similarly, a meta-analysis by Wu, Chiou, Kao, Hu, and Huang (2012) found that only 91 of 567 studies were based on learning theory, with most being descriptive in nature and none being classified within a strictly SRL framework. However, there is promise for the study of SRL in gaming; Vogel et al. (2006) conducted a meta-analysis and found that computer games and interactive simulations where learners controlled the navigation resulted in higher cognitive gains compared with traditional teaching methods. In addition, SRL training as a precursor to learning sessions and externally mediated human tutoring in hypermedia environments have shown promise for increasing content understanding and mental model development (Azevedo & Cromley, 2004; Johnson, Azevedo, & D'Mello, 2011).

Although the gender gap in academic achievement has narrowed in recent years (Wirt et al., 2004), research has shown important differences between male and female students in how they regulate their learning (Arroyo, Burlison, Tai, Muldner, & Woolf, 2013). Arroyo et al. (2013) recommended customizing technology-based learning by gender after finding differences in help seeking and the ability to work with animated learning companions in a mathematics learning environment for middle and high school students. In the current study, our examination of gender was limited to SRL variables that (a) prior research would suggest to be critical for performance within educational games and (b) could be operationalized within the context of Crystal Island– Outbreak.

Strategy Use

Gender and cognitive strategy use have been studied in traditional educational settings and to a lesser extent in game-based learning environments. Studies have shown that as early as first grade, gender differences in the use of cognitive strategies for math is apparent (Carr, Jessup, & Fuller, 1999). In game-based learning environments, effective strategy use is critical to performance and can be operationalized as such behaviors as effective tool use, time management, and appropriate help seeking, just to name a few. Recently, methodological advances have gone beyond self-report and instead encourage the use of trace data in computer-based learning environments and programs that employ scaffolded note-taking functionalities (Winne & Hadwin, 2013). In game-based learning environments, middle school girls have been shown to

record more notes than boys (McQuiggan, Goth, Ha, Rowe, & Lester, 2008). Girls have also reported a preference for exploration in such environments, whereas boys have reported more strategic approaches for using in-game resources (Hamlen, 2011; Kinzie & Joseph, 2008).

Metacognition

Metacognition encompasses knowledge and regulation of one's online and offline cognitive processes (Schraw & Moshman, 1995). Girls have been shown to report higher levels of metacognitive strategy use, such as record keeping and monitoring, across grade levels (Grades 5, 8, and 11) than boys (Zimmerman & Martinez-Pons, 1990). Metacognitive monitoring and control processes are critical in game-based learning environments in order to adapt to changing narratives, judge the effectiveness of chosen strategies, and choose efficient problem-solving pathways. In the current study, we were interested in comparing online metacognitive monitoring judgments with actual performance. These types of judgments have been measured using confidence estimates to gauge calibration, or the match between one's perception of performance and that person's actual level of performance (Dunlosky & Metcalfe, 2009; Nietfeld, Cao, & Osborne, 2006). Studies have shown that males tend to be more confident than females when making academic global judgments of confidence; however, less is known about the role of gender for item-specific confidence that indicates calibration (Lundeberg & Mohan, 2009). The examination of calibration and gender related to differences in calibration within educational games has been minimal to date.

Motivation

We limited our focus within motivation in the current study to goal orientation, interest, and self-efficacy. These motivational constructs have been prevalent across empirical studies and models of SRL (Pintrich, 2000; Winne & Hadwin, 1998; Zimmerman, 2000) and are directly relevant to instruction in science-focused games.

Goal orientation is an individual difference construct that impacts academic motivation and classroom achievement (Elliot & Dweck, 1988). Two primary goal orientations have dominated the literature: mastery (or learning) orientation and performance orientation (Ames, 1992; Dweck, 1986). Mastery-oriented individuals focus on gaining knowledge, increasing ability, and mastering a certain amount of material. These individuals tend to be more persistent (Dweck, 1986) and report more feelings of regulatory control (Roedel, Schraw, & Plake, 1994). Performance-oriented individuals seek to outperform their peers (Elliot & McGregor, 2001). Empirical evidence is lacking for the role of goal orientations in educational games; however, one might posit they play a role in determining the emphasis one has in achieving points and in-game rewards versus gaining content mastery while also influencing the extent to which one persists in attempting to solve a complex problem. For both genders, mastery goals have been associated with higher levels of cognitive strategy use (Wolters, Yu, & Pintrich, 1996), metacognitive strategy use (Wolters, 2004), and interest (Wolters et al., 1996). Studies of elementary-age science students have shown mixed results with regard to gender differences for mastery goal orientation (Anderman & Young, 1994; Meece & Painter, 2008).

Interest has been defined as a motivational belief (Zimmerman, 2002) and a construct comprising both cognition and affect that leads to engaged learning (Hidi & Renninger, 2006). Interest can be considered as either personal or situational (Schraw & Lehman, 2001). Personal interest is enduring and context-general, whereas situational interest is spontaneous and context-specific. Studies in science have shown that students who report enjoyment in solving problems are more likely to report a desire to engage with similar content in the future (Ainley & Ainley, 2011). Typically, educational games seek to capture students' situational interest with the intent that it may eventually lead to more enduring personal interest. Within video games, girls tend to show more situational interest in story development, relationships, and collaboration, whereas boys tend to prefer competition and aggression (Cassell & Jenkins, 1998). Eighth-grade girls have been found to have higher interest than boys in more traditional, thoughtful games, whereas eighth-grade boys prefer physical and imagination games more so than girls (Greenberg, Sherry, Lachlan, Lucas, & Holmstrom, 2010).

Self-efficacy has been defined as one's confidence in his or her ability to plan and implement a course of action to solve a problem or accomplish a task (Bandura, 1997) and has been hypothesized to influence an individual's choice, effort, and persistence in an activity (Bandura, 1986). Crystal Island–Outbreak was designed to engage students in a challenging, mastery-oriented problem-solving context, thereby supporting work in self-efficacy demonstrating the impact of mastery experiences on self-efficacy (Britner & Pajares, 2006). Gender-related research in self-efficacy has found that boys report higher levels of self-efficacy for science than girls in elementary school (Meece & Jones, 1996), but findings for middle school have been mixed (Anderman & Young, 1994; Britner & Pajares, 2006). In computer-related judgments of self-efficacy, boys have been shown to have clear advantages (Durdell & Haag, 2002).

Crystal Island–Outbreak

Crystal Island–Outbreak is an immersive, three-dimensional (3D), inquiry-based game-based learning environment in the domain of microbiology created for eighth-grade students. The program was implemented using Valve's Software Source Engine, the 3D game platform for Half-Life 2. The game features a science mystery set on a recently discovered volcanic island where a research station has been established to study the island's unique flora and fauna. The user plays the protagonist, Alyx, a gender-neutral character, who is attempting to find the source of the disease that is making other researchers sick. As research team members fall ill, the learner must discover the specific source of the outbreak by exploring the world and interacting with characters who are experts on various microbiology topics while forming questions, generating hypotheses, collecting data, and testing hypotheses. In order to "win" the game, the student must complete and submit a correct diagnosis with information about the source, disease, and treatment. A visible running score was provided to the students during game play that involved rewarding students for completing mini-goals in an efficient manner (based on latency).

The creation of Crystal Island–Outbreak has been guided theoretically by a combination of perspectives on learning and engagement. These include need for structure (Kirschner, Sweller, &

Clark, 2006) in order to build rich, interconnected background knowledge and also constructivist approaches that provide learners with an opportunity to self-regulate their problem-solving approaches with appropriate scaffolding. Positive academic outcomes are more likely to arise when students show motivational indicators of engagement, such as being fully immersed, interested, and actively participating (Skinner & Belmont, 1993; Skinner, Furrer, Marchand, & Kindermann, 2008). In addition, cognitive engagement stresses the volitional investment (Corno, 1993) in learning and the ability to be strategic and to self-regulate (Fredricks, Blumenfeld, & Paris, 2004). Crystal Island–Outbreak and similar game-based environments have been shown to elicit engagement or flow as the most frequently occurring affective state (D'Mello, 2013), a finding important given engagement's role as a precursor to cognitive and metacognitive processes that aid in SRL. A challenge for the current study, as in other similar studies (Winne & Perry, 2000), was in the measurement of SRL variables within Crystal Island–Outbreak. Thus, we sought to include measures consistent with current taxonomies within the field that included both offline and online measures (Schraw, 2010). These included measures of cognitive abilities (knowledge and strategy use), self-reported beliefs, trace logs, and measures of calibration.

A number of changes from previously tested versions (Adams, Mayer, MacNamara, Koenig, & Wainess, 2012) of Crystal Island–Outbreak were made to enhance the learning experience. A structured note-taking tool called the *diagnosis worksheet* was introduced to assist the learner in tracking and organizing important information and to measure cognitive strategy use. Communication was enhanced to include voice-acted spoken dialog, which simultaneously appeared in text format across the bottom of the screen. The game also included a personal digital assistant (PDA) device that allowed characters to communicate with the students and provided microbiology-based content questions throughout game play.

Current Study

The goal of this study was to investigate the influence of self-regulation and gender within Crystal Island–Outbreak. Two primary research questions were examined: First, *what SRL variables predict performance in Crystal Island–Outbreak?* This question was important in order to understand whether performance in game-based learning environments such as Crystal Island–Outbreak is simply due to prior knowledge and gaming history or would require effective self-regulation for successful performance. We carefully chose SRL variables for this study previously shown to significantly impact academic learning. Thus, we predicted that each variable would contribute significant variance to in-game performance within Crystal Island–Outbreak over and above variance accounted for by prior knowledge and perceived gaming skill.

The use of in-game performance as a dependent measure reflects the efficiency of solving an applied scientific problem (Linn & Eylon, 2006). In essence, Crystal Island–Outbreak presents a customized inquiry-based problem-solving exercise in an immersive digital environment that requires focus, efficiency, and mastery of content knowledge in order to perform well. The game targets basic microbiology content knowledge as well as skills such as deductive reasoning, critical thinking, and the ability to

efficiently carry out problem-solving steps. Therefore, students are required to make judgments about what information is more versus less important for solving the mystery, understand which sources of information (characters, texts, posters) are credible, and learn how to navigate the environment efficiently in the face of environmental distractions. Prior studies that have tested Crystal Island–Outbreak with students in the target age group (eighth graders) have shown it to be effective at increasing content knowledge (Nietfeld, Shores, & Hoffmann, 2011; Rowe et al., 2010). We argue that these findings plus improvements to the game itself provide a meaningful justification for using in-game performance as an outcome measure of scientific problem solving. Upon starting the game, students are aware of the overall goal and must successfully navigate through a number of subgoals. Therefore, the skills required to solve the mystery in Crystal Island–Outbreak are akin to those required to solve a complex performance-based problem in a lab-based setting.

The second research question was *What role does gender play with regard to SRL in Crystal Island–Outbreak?* Given the attention that gender has received within the educational computer games literature (Carr & Pelletier, 2009; Hayes, 2011; Plant et al., 2009), it is important that such environments are effective and efficient in assisting students of both genders to develop SRL skills. With regard to cognitive strategy use, we compared boys and girls on their use of the diagnosis worksheet. We predicted higher levels of in-game performance for boys on the diagnostic worksheet, given prior research showing their advantages in strategic approaches using in-game resources (Hamlen, 2011; Kinzie & Joseph, 2008). For metacognition, we compared monitoring judgments for content-related questions spread throughout gameplay. We expected females to be more accurate and less overconfident on these judgments; however, the lack of research on monitoring judgments in similar environments made this prediction somewhat tenuous. Motivation comparisons were made between genders for goal orientation, self-efficacy, and situational interest using self-report inventories. Given the lack of gender differences in goal orientation found within traditional environments (Meece & Painter, 2008) and the mixed findings for self-efficacy in middle school (Anderman & Young, 1994; Britner & Pajares, 2006), we predicted similar predictive patterns for in-game performance across genders. However, findings in favor of boys for experience in digital gaming environments (Greenberg et al., 2010) led us to predict that situational interest would be a stronger predictor of in-game performance for boys than girls.

Method

Participants

A total of 130 eighth-grade students (73 boys, 57 girls) from a middle school in the Southeast who ranged in age from 12 to 15 years ($M = 13.28$, $SD = 0.48$) participated in this study. Racial and ethnic breakdown of the participants was as follows: 2.3% American Indian/Native American, 1.5% Asian/Asian American, 30.8% African American, 12.3% Hispanic, and 46.9% White; 6.2% did not report race or ethnicity. Approximately 38% of the students were eligible for free/reduced-price lunches, and the school performed at average levels relative to other schools in the state with regard to overall academic performance based upon yearly state tests. The study was

conducted prior to students' exposure to microbiology curriculum in their regular classes.

Measures

Gaming survey. A short gaming survey was created as a pretest for this study that contained questions regarding experience with video games and computer use. We were particularly interested in two 5-point Likert scale items that asked, "How skilled are you when playing video games?" to assess perceived skill and "How frequently do you play video games?" to assess experience.

Microbiology content knowledge test. This test consisted of 16 multiple-choice items based on both the state's standard course of study for eighth-grade science and Crystal Island–Outbreak itself. Eight of the items were fact level while eight of the items were application level. Content validity was assured through alignment with the state's standard course of study and through a teacher and science educator review process. The test was administered as a measure of prior knowledge ($\alpha = .16$) and again as a posttest ($\alpha = .61$) to measure content learning gains. Internal reliability was extremely low for the pretest, most likely because students had not been exposed to the content and had very little background knowledge or experience to draw upon when answering the items. No feedback on item correctness was provided during either administration of the test. The following excerpts are example items:

Fact-level item: A pathogen is a microbe that:

- (a) has one specialized route of transmission.
- (b) causes disease in another organism.
- (c) is not capable of spreading from organism to organism.
- (d) does not require a host in order to reproduce.

Application-level item: Your lab partners ask you to help them identify a pathogen. Unfortunately, they do not know the size of the organism, but they know that it is smooth and round. What might your lab partners be looking at?

- (a) A virus.
- (b) A bacterium.
- (c) A fungi.
- (d) Both a and b.

During the student's interactions with expert characters in Crystal Island–Outbreak, in-game, multiple-choice items were presented to test the student's knowledge of that character's specific expertise. There were a total of 18 possible items, and each question was followed by a prompt to assess the student's confidence in his or her answer. The confidence question asked the student to "indicate how confident you are that you answered the previous question correctly. 0 represents a complete lack of confidence, 100 represents total confidence, and 1–99 represent different degrees of confidence." The questions were presented through the character as an extension of their dialogue so as not to significantly interrupt the flow of the game. All students did not receive all 18 questions because of variation in gameplay at the individual level.

Strategy use. The diagnosis worksheet (see Figure 1) created for Crystal Island–Outbreak functioned as a cognitive tool and structured note-taking scaffold in which the students recorded information via drop-down boxes rather than taking free-form notes. The worksheet scaffold was divided into four distinct sections: one for recording patient symptoms (8 points), one for

DIAGNOSIS WORKSHEET

Patients' Symptoms		Test Results	
Symptom 1:	Vomiting	I have tested:	The lab results showed:
Symptom 2:	Pain	Object 1:	Orange
Symptom 3:	Pain	Object 2:	No Entry
Symptom 4:	No Entry	Object 3:	No Entry
		Object 4:	No Entry
			Not contaminated
			No Entry
			No Entry
			No Entry

HYPOTHESES

Select possible diagnosis:	Likelihood of this diagnosis:	Because:
Influenza	Unlikely	Symptoms don't match
Anthrax	Very Unlikely	Characteristics don't match
No Entry	No Entry	No Entry

FINAL DIAGNOSIS

I believe the illness is:	No Entry	I believe the appropriate treatment is:	No Entry
I believe the transmission source is:	No Entry	because the type of infection is:	No Entry

Figure 1. The Crystal Island–Outbreak diagnosis worksheet.

collecting information regarding contaminated objects after testing them in the laboratory (16 points), one for constructing a well-formed and justified hypothesis (28 points), and one for determining a final diagnosis (48 points). A point system was calculated based upon the thoroughness and accuracy of the use of the worksheet. Sections were weighted according to the amount of student-generated analysis that was required. For example, the symptoms section simply required students to record the symptoms expressed by the sick patient characters; however, the hypothesis section required students to analyze their collected information and to generate their own hypotheses for the nature and likelihood of the illness. This point system was then interpreted as a measure of in-game strategy use.

Metacognitive judgments. The confidence estimates provided by the in-game quiz questions allowed for the calculation of two metacognitive monitoring scores that included a *calibration* score and a *bias* score (Schraw, 1994). Calibration represented the absolute difference in total confidence scores minus percentage of accuracy on the quiz questions. Bias measured the signed difference in total confidence scores minus percentage correct on quiz questions. Thus, if a student had an average confidence judgment of 79 (.79) and correctly answered 84% of the questions (.84), their accuracy score would be .05 and their bias score would be $-.05$. For bias, positive scores represent overconfidence and negative scores underconfidence. Monitoring scores were calculated and averaged across all items to arrive at a single index of calibration and bias. The two indices provide complementary information as

calibration reports to the extent to which one's judgment matches actual performance while bias indicates the degree to which an individual is generally over- or underconfident.

Transfer. Two open-ended tasks measuring near transfer were also included after gameplay.

Near Transfer Task 1: "Imagine that you have three microbes that are three different sizes. Please explain how you could identify each microbe if you know that one is a virus, one is a bacterium, and one is a fungus."

Near Transfer Task 2: "A scientist wonders if a new microbe she has found could cause illness in humans. She wants to be a good scientist and has come to you for advice. In this specific situation, develop a set of instructions to complete each step of the scientific method."

The items required students to apply problem-solving processes in a context outside the game requiring a deeper level of understanding than identification-level multiple-choice items. Responses to these items were scored on a 3-point scale for accuracy and completeness of the answer and coded separately by two raters who came to agreement on all scoring discrepancies after discussion.

Motivational measures.

Achievement Goals Questionnaire (AGQ) Scale. Goal orientation was measured as a pretest using the 12-item AGO scale that measures achievement goal orientation in the form of four factors: mastery approach, mastery avoidance, performance approach, and

performance avoidance (Elliot & McGregor, 2001). The four factors have been shown to have strong reliability with internal reliabilities all over .70 and evidence of discriminant validity with low-to-moderate correlations among the four goal orientations (Finney, Pieper, & Barron, 2004). Coefficient alphas for the mastery approach and performance approach facets in the current study were .77 and .80, respectively.

Science Learning Self-Efficacy Inventory. This eight-item Likert-scale inventory was also given prior to the students' interaction with Crystal Island–Outbreak. The inventory was adapted from Nietfeld et al. (2006) where the scale was shown to be internally reliable ($\alpha = .88$ and $.90$) and predictive of academic performance. A sample item included, "I think I have good skills and strategies to learn science." The items deviated somewhat from most contemporary self-efficacy scale items that are very task specific (Bandura, 1997) and instead referred to confidence in one's ability to perform well and understand topics more generally in science. Coefficient alpha in this study was .84, and correlations with prior knowledge ($r = .24$) and posttest knowledge of microbiology ($r = .20$) were significant ($p < .05$).

Perceived Interest Questionnaire (PIQ). At the end of the Crystal Island–Outbreak interaction, students were asked to complete the 10-item PIQ adapted from Schraw (1997), which measured their situational interest in the game ("Did you enjoy this game?"). Schraw (1997) used factor analysis to determine that the items on the PIQ measured a single factor of interest and also had strong internal reliability ($\alpha = .92$). Students responded to the PIQ items ($\alpha = .94$) on a 5-point Likert scale, with 1 being *Not at all true of me* and 5 being *Very true of me*.

In-game performance. Computation of the score in Crystal Island–Outbreak involved rewarding students for completing goals (mini-milestones working toward the solution) and for efficiency (based on latency). Students could also make a guess at the correct solution at any point in the game, resulting in a point benefit for accuracy and a point loss for inaccuracy. In-game actions that were scored included conversing with nonplayer characters, performing well on the in-game multiple-choice items presented by characters, testing objects for contaminants, a pathogen-labeling activity, and submitting a correct diagnosis worksheet.

Procedure

Participants completed the gaming survey, self-efficacy survey, and the AGQ approximately 1 week prior to playing Crystal Island–Outbreak (see Figure 2). On the day of the experiment,

students were provided with general details about the Crystal Island–Outbreak mystery and game controls during an introductory presentation by a researcher. After the presentation, students completed the test of content knowledge for microbiology. The students were then given 60 min to work individually in a computer lab on solving the mystery in Crystal Island–Outbreak. The postexperiment materials included the posttest of content knowledge, transfer tasks, and measure of situational interest and were completed immediately after gameplay.

Results

Descriptive statistics for the major study variables are presented in Table 1 and bivariate correlations are provided in Table 2. Overall comparisons of microbiology prior knowledge ($M = 6.26$, $SD = 1.99$) and microbiology posttest knowledge ($M = 8.66$, $SD = 2.98$) revealed a significant increase in scores, $t(129) = 9.81$, $p < .001$; $d = 0.91$. In addition, a significant relationship ($r = .27$; $p = .002$) was found between learning gains on the content knowledge posttest and score in Crystal Island–Outbreak. Score in Crystal Island–Outbreak was also correlated with both of the open-ended transfer tasks ($r = .23$; $p = .007$, and $r = .25$; $p = .005$, respectively).

Although causality could not be established in the current design, comparisons were made between students who completed the Crystal Island–Outbreak mystery and those who did not for variables related to SRL and content knowledge. Overall, 58 students completed the mystery. An analysis of covariance (ANCOVA) controlling for prior knowledge, $F(1, 127) = 18.33$, $p < .001$, partial $\eta^2 = .13$, found students who completed the Crystal Island–Outbreak mystery scored significantly better on the microbiology posttest than students who did not complete the scenario, $F(1, 127) = 20.12$, $p < .001$, partial $\eta^2 = .14$; assumption of homogeneity of regression slopes for covariate maintained, $F = 1.32$, $p = .25$. Moreover, students who solved the mystery reported significantly higher levels of situational interest, $M_{(\text{completed})} = 33.97$, $SD = 7.49$, vs. $M_{(\text{not completed})} = 30.36$, $SD = 8.99$; $t(128) = 2.45$, $p < .05$; and self-efficacy for science, $M_{(\text{completed})} = 32.21$, $SD = 5.98$, vs. $M_{(\text{not completed})} = 28.11$, $SD = 4.60$; $t(128) = 4.41$, $p < .001$. They also had a higher diagnosis worksheet total, $M_{(\text{completed})} = 78.53$, $SD = 14.69$, vs. $M_{(\text{not completed})} = 25.96$, $SD = 24.97$; $t(128) = 14.14$, $p < .001$; were less overconfident, $M_{(\text{completed})} = -0.05$, $SD = 0.30$, vs. $M_{(\text{not completed})} = 0.06$, $SD = 0.30$; $t(128) = 2.06$, $p < .05$; and performed better on both near transfer problems—Transfer 1 = $M_{(\text{completed})} = 0.90$, $SD = 0.77$,

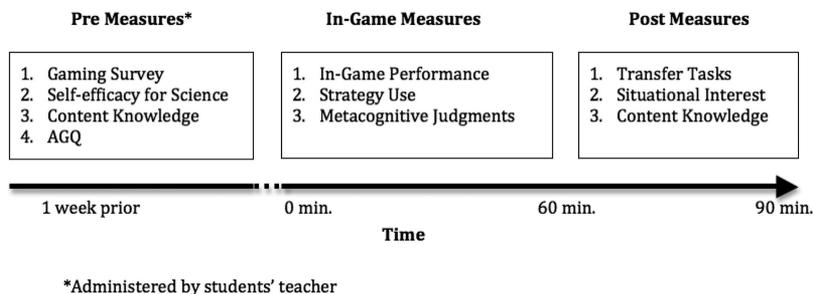


Figure 2. Summary of Study 2 procedures. AGQ = Achievement Goals Questionnaire.

Table 1
Means (and Standard Deviations) of Study Variables

Variable	Boys	Girls	Total
Prior knowledge	6.33 (2.21)	6.18 (1.70)	6.26 (1.99)
Posttest content	8.79 (3.00)	8.49 (2.98)	8.66 (2.98)
Perceived game skill	4.11 (0.94)***	2.79 (1.09)	3.53 (1.20)
Game play frequency	3.79 (1.01)***	2.51 (0.78)	3.23 (1.12)
Diagnosis worksheet use	54.90 (32.63)*	42.39 (33.73)	49.42 (33.57)
Goals completed	6.07 (0.98)**	5.56 (1.21)	5.85 (1.11)
Total score	462.21 (442.19)	355.09 (399.79)	415.24 (425.87)
Calibration	0.26 (0.18)	0.25 (.17)	0.25 (0.17)
Response bias	0.03 (0.31)	0.00 (.30)	0.01 (0.31)
Self-efficacy for science	30.14 (5.98)	29.68 (5.17)	29.94 (5.63)
Mastery approach	17.44 (3.44)	18.02 (2.88)	17.69 (3.21)
Mastery avoid	10.97 (4.88)	11.04 (4.90)	11.00 (4.87)
Performance approach	13.42 (4.69)	14.16 (4.85)	13.75 (4.76)
Performance avoid	15.08 (4.78)	16.49 (4.03)	15.70 (4.51)
Interest	32.33 (8.31)	31.51 (8.81)	31.97 (8.51)

* $p < .05$. ** $p < .01$. *** $p < .001$.

vs. $M_{(\text{not completed})} = 0.58, SD = 0.67; t(128) = 2.49, p < .05$; and Transfer 2 = $M_{(\text{completed})} = 0.72, SD = 0.77$, vs. $M_{(\text{not completed})} = 0.36, SD = 0.63; t(128) = 2.95, p < .01$. Those who completed the mystery did not report higher levels of gaming skill, $t(128) = 0.47, p = .64$, or gameplay frequency, $t(127) = 0.65, p = .52$.

What SRL Variables Predict Performance in Crystal Island–Outbreak?

In order to test the contribution of SRL on performance (game score) in Crystal Island–Outbreak, we conducted a regression analysis (see Table 3). The variables were carefully chosen to represent each of the three major facets of SRL (strategy use, metacognition, motivation). These included the diagnosis worksheet score (strategy use), monitoring bias from in-game content items (metacognition), self-efficacy for science (motivation), situational interest (motivation), and mastery-approach and performance-approach goal orientations (motivation). In addition, it was important to include prior knowledge and perceived gaming skill in order to control for pre-existing differences between students. The eight variables were entered simultaneously. The model significantly predicted game score,

$F(8,121) = 10.95, R^2 = .420, p < .001$. Five of the variables were found to be significant predictors, and they included diagnosis worksheet score ($\beta = .362, p < .001$), situational interest ($\beta = .199, p = .007$), self-efficacy ($\beta = .224, p = .007$), monitoring bias ($\beta = -.256, p = .001$), and prior knowledge ($\beta = .156, p = .039$). Perceived gaming skill, mastery approach, and performance approach were not significant predictors.

What Role Does Gender Play With Regard to SRL in Crystal Island–Outbreak?

In this section, we will first report on performance in Crystal Island–Outbreak with regard to gender followed by an analysis of SRL facets including strategy use, metacognition, and motivation. Prior to interacting with Crystal Island–Outbreak, boys (frequency: $M = 3.79, SD = 1.01$; skill: $M = 4.11, SD = .94$) reported playing games more frequently, $t(128) = 7.91, p < .001; d = 1.43$, and also reported higher levels of perceived gameplay skill, $t(128) = 7.39, p < .001; d = 1.29$, than girls (frequency: $M = 2.51, SD = 0.78$; skill: $M = 2.79, SD = 1.10$). A hierarchical regression analysis was conducted using gender and reported number of gameplay hours to predict in-game goals completed. Gender was

Table 2
Bivariate Correlations Among Major Study Variables

Measure	1	2	3	4	5	6	7	8	9	10	11	12
1. Prior knowledge	—	.43**	.13	.08	.29**	.30**	-.11	.14	.24**	.03	.08	.15
2. Posttest knowledge		—	-.05	.05	.41**	.45**	-.03	-.11	.20*	.05	-.03	.19*
3. Game skill			—	.37**	.07	.00	-.19*	.08	.09	.08	-.07	.07
4. Game frequency				—	.15	.11	-.03	.10	-.02	-.16	-.19*	.14
5. Diagnosis worksheet use					—	.73**	-.09	-.19*	.36**	-.02	.07	.12
6. Game score						—	-.04	-.22*	.37**	-.01	.12	.19*
7. Calibration							—	-.02	-.08	-.08	.00	-.19*
8. Response bias								—	.14	-.08	-.07	.17*
9. Science self-efficacy									—	.20*	.26**	.12
10. Mastery approach										—	.31**	.22*
11. Performance approach											—	.05
12. Interest												—

* $p < .05$. ** $p < .01$.

Table 3
Regression Results Predicting Performance in Crystal Island–Outbreak

Predictors	B	SE	β
Constant	-550.24	235.37	
Diagnosis worksheet use	10.48	2.31	.36***
Interest	9.97	3.66	.20**
Self-Efficacy for science	16.95	6.14	.22**
Monitoring bias	-355.37	103.93	-.26**
Prior knowledge	33.33	15.96	.16*
Perceived game skill	-20.11	25.09	-.06
Mastery approach	-16.02	10.15	-.12
Performance approach	3.45	6.74	.04
R^2		.420	

Note. Dependent variable is performance in Crystal Island–Outbreak. SE = standard error.

* $p < .05$. ** $p < .01$. *** $p < .001$.

entered into the first block, and reported number of gameplay hours was added to the second block. Both models were found to be significant—Block 1: $\beta_{\text{gender}} = -.22$, $R^2 = .05$, $F(1, 127) = 6.24$, $p < .05$; Block 2: $\Delta R^2 = .02$, $F(2, 126) = 4.56$, $p < .05$. However, when adding reported number of gameplay hours in the second block, gender was no longer a significant predictor of goals completed. Thus, when we controlled for number of reported hours playing video games, boys were no longer advantaged in terms of goals completed. Also, boys tended to receive significantly higher in-game scores than girls toward the beginning of the interaction, but these differences dissipated to nonsignificant group differences by the end of gameplay. Significant differences in favor of boys ($M_{\text{boys}} = 34.78$, $SD = 24.42$; $M_{\text{girls}} = 23.26$, $SD = 18.86$) were found after 5 min, $t(128) = 2.94$, $p < .01$, and 25 min ($M_{\text{boys}} = 210.38$, $SD = 239.39$; $M_{\text{girls}} = 135.07$, $SD = 115.71$), $t(128) = 2.18$, $p = .03$, but not after 45 min ($M_{\text{boys}} = 378.89$, $SD = 413.54$; $M_{\text{girls}} = 278.51$, $SD = 329.26$), $t(128) = 1.50$, $p = .14$, or 60 min ($M_{\text{boys}} = 462.21$, $SD = 442.19$; $M_{\text{girls}} = 355.09$, $SD = 399.79$), $t(128) = 1.43$, $p = .16$. No significant differences were found between genders for prior knowledge or posttest performance on microbiology. However, both genders showed significant increases in content knowledge (for boys, $M_{\text{pretest}} = 6.33$, $SD_{\text{pretest}} = 2.21$; $M_{\text{posttest}} = 8.79$; $SD_{\text{posttest}} = 3.00$), $t(72) = -7.76$, $p < .001$, $d = 0.94$; (for girls, $M_{\text{pretest}} = 6.18$, $SD_{\text{pretest}} = 1.70$; $M_{\text{posttest}} = 8.49$; $SD_{\text{posttest}} = 2.98$), $t(56) = -6.02$, $p < .001$, $d = 0.85$. Finally, boys showed a significant relationship ($r = .35$) between their prior knowledge for microbiology and their in-game score, but no significant relationship was found with girls.

Of the 73 boys who interacted with Crystal Island–Outbreak, 36 completed the mystery. Within the sample of boys, those who completed the mystery had significantly higher levels of prior knowledge, $t(71) = 3.05$, $p < .01$, $d = 0.72$; situational interest, $t(71) = 2.26$, $p < .05$, $d = 0.54$; self-efficacy for science, $t(71) = 3.30$, $p < .01$, $d = 0.78$; diagnosis worksheet total, $t(71) = 9.74$, $p < .001$, $d = 2.50$; performance on the microbiology posttest, $t(71) = 4.41$, $p < .001$, $d = 1.04$; and performance on the transfer tasks, $t(71) = 2.45$, $p < .01$, $d = 0.57$; and $t(71) = 3.06$, $p < .01$, $d = 0.73$, respectively.

Of the 57 girls, 22 completed the Crystal Island–Outbreak mystery. Considering only the female sample, those who com-

pleted the mystery reported significantly higher levels of self-efficacy for science, $t(55) = 2.84$, $p < .01$, $d = 0.76$; diagnosis worksheet performance, $t(55) = 10.25$, $p < .001$, $d = 2.80$; and performance on the microbiology posttest, $t(55) = 3.28$, $p < .01$, $d = 0.95$. No significant differences in microbiology prior knowledge were found between girls who completed the mystery and girls who did not complete the mystery.

Strategy use. Boys ($M = 54.90$, $SD = 32.63$) more accurately and comprehensively completed the diagnosis worksheet, $t(128) = 2.14$, $p < .05$, $d = 0.38$, than girls ($M = 43.39$, $SD = 33.73$). Boys scored higher (mandatory diagnosis: $M = 30.05$, $SD = 20.17$; testing: $M = 11.19$, $SD = 7.91$) than girls (mandatory diagnosis: $M = 21.47$, $SD = 21.60$; testing: $M = 8.46$, $SD = 7.39$) on the mandatory diagnosis section, $t(128) = 2.33$, $p < .05$, $d = 0.41$, and the testing section, $t(128) = 2.01$, $p < .05$, $d = 0.36$, of the diagnosis worksheet but not on the hypothesis or symptoms section. However, gender differences in diagnosis worksheet performance disappeared when controlling for number of hours reported playing video games.

Metacognition. Overall, no significant differences were found between genders in monitoring ability for either calibration or bias on the in-game quiz questions. Averaged across students, neither gender (boys, $M = 0.02$; girls, $M = 0.00$) showed much response bias (over- or underconfidence); however, there was significant variance within groups as evidenced by the standard deviations (boys = 0.31; girls = 0.30). Calibration rates revealed that the average confidence judgment for boys was off by 26 points (out of 100) from their average overall percentage of correct across items compared with 25 points for girls. Interestingly, our data suggest that for boys it was beneficial to be underconfident rather than overconfident, but this did not hold for girls. In order to examine this finding more closely, we converted scores for monitoring bias to standardized z scores and used them to create three groups, splitting the sample into approximate thirds. The *underconfident* group (boys $N = 21$; girls $N = 18$) consisted of all students who had a z score of $-.5$ or less. The *accurate* group (boys $N = 28$; girls $N = 20$) consisted of all students who had a z score between $-.49$ and $.49$. The *overconfident* group (boys $N = 24$; girls $N = 19$) consisted of all students who had a z score higher than $.5$. We then conducted separate ANCOVAs by gender to examine the impact of confidence level on Crystal Island–Outbreak score and microbiology posttest content knowledge, with prior knowledge included as a covariate (see Figure 3). Boys in the underconfident group ($M = 9.57$, $SD = 3.82$, adjusted $M = 10.02$, standard error [SE] = 0.56) scored significantly higher, $F(2, 69) = 3.43$, $p = .038$, $\eta^2 = .09$, assumption of homogeneity of regression slopes for covariate maintained, $F = 0.59$, $p = .56$, on posttest microbiology content knowledge than their peers in the overconfident group ($M = 8.33$, $SD = 2.51$, adjusted $M = 8.12$, $SE = 0.52$). This finding was replicated for Crystal Island–Outbreak game score, $F(2, 69) = 3.60$, $p = .032$, $\eta^2 = .10$, assumption of homogeneity of regression slopes for covariate maintained, $F = 1.16$, $p = .32$, as the underconfident group (observed $M = 572.19$, $SD = 503.31$, adjusted $M = 623.34$, $SE = 88.13$) outperformed the overconfident group ($M = 325.75$, $SD = 397.50$, adjusted $M = 301.09$, $SE = 81.74$). Analysis of bias scores for girls failed to show any significant differences by groups for these two variables.

Motivation. In order to analyze the predictive strength of the motivational variables measured in this study with regard to in-

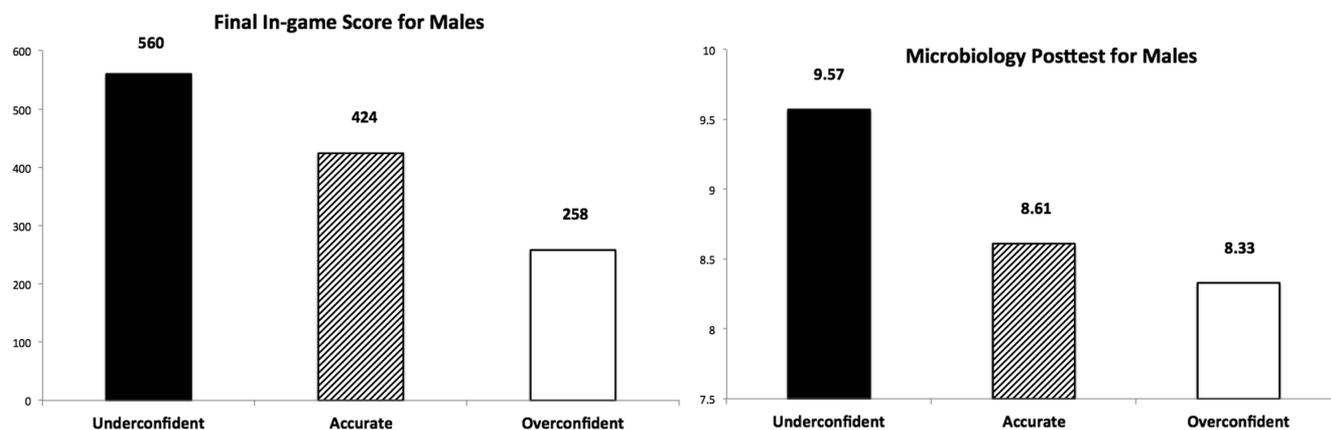


Figure 3. The impact of response bias for boys on in-game score and the microbiology posttest.

game performance, we conducted separate multiple regression analyses by gender. For each analysis, science self-efficacy, situational interest, performance approach, and mastery approach were entered as predictors of in-game score. Results revealed that only the model for boys was significant, $F(8, 121) = 4.13$, $R^2 = .195$, $p = .005$. Within that model, only science self-efficacy was a significant predictor ($\beta = .364$, $p = .003$) while situational interest approached significance ($\beta = .194$, $p = .083$). Additionally, for girls, there was a significant bivariate correlation between mastery approach scores and interest ($r = .37$), but no significant relationship was found for boys. No significant differences were found for levels of situational interest between boys and girls. Finally, no significant correlations were found between mastery approach and scores on the diagnosis worksheet for either gender.

Discussion

Findings and Implications

A growing literature is showing promise for student content learning in educational games (Connolly et al., 2012; Wouters et al., 2013). Yet, little is known about how SRL and gender influence and predict performance in these environments. In the current study, we sought to investigate these individual difference variables in Crystal Island–Outbreak, a game-based learning environment representative of other educational games in that its content (microbiology) is aligned with school curriculum and attempts to scaffold learning in an engaging setting. Our data provided a number of interesting findings related to the impact of SRL and gender in predicting performance while also supporting Crystal Island–Outbreak as effective for increasing content knowledge. Three of the major findings and their broader implications are described in detail in the following text.

First, in-game performance in Crystal Island–Outbreak was predicted by SRL variables that included effective strategy use (as evidenced through the use of the diagnosis worksheet), metacognitive monitoring, and motivational variables such as situational interest and self-efficacy. Our hypothesis for this analysis was largely supported, with the exception being that goal orientations were not significant predictors. This finding is important in that evidence in the literature examining the independent contribution

of variables from all three major SRL facets in educational games is scarce. Students who solved the mystery within Crystal Island–Outbreak showed similar relationships as they had higher self-efficacy and situational interest and displayed more effective strategy use (diagnosis worksheet) as well as having more prior knowledge than their peers who did not solve the mystery. The relationship between the quality of the diagnosis worksheet use and in-game performance is likely due to the fact that it facilitates the tracking, organizing, and off-loading of incoming information, all behaviors indicative of effective SRL. Our findings also suggest that overconfidence may hinder both performance and self-regulation, particularly for boys, within educational games similar to Crystal Island–Outbreak.

Open-ended or “realistic” environments that offer choice and autonomy are necessary both for the development of effective SRL and the observation of SRL unfolding (Schunk & Zimmerman, 2003). In accordance with this statement, our data suggest that SRL skills were essential components to successful performance in Crystal Island–Outbreak. It appears that we achieved some level of success in presenting a learning environment that elicited indicators of student engagement (Skinner et al., 2008), maintained student focus (Fredricks et al., 2004), and appealed to both genders (Cassell & Jenkins, 1998; Greenberg et al., 2010) in order to facilitate problem solving and content learning. Future researchers should consider employing games with even more complex problem scenarios that are integrated in classroom-based settings for longer periods of time (Jackson & McNamara, 2013). Such designs that have students going in and out of the game while also engaging in more traditional content-related classroom activities have the potential to have a greater impact on SRL skills and potentially for transfer of concepts outside the game itself.

Where do findings in the current study leave us with regard to SRL in game-based learning environments? Given that we found measures representing all three major dimensions of SRL to predict in-game performance, it will be important to examine components of educational games that are effective at initiating self-regulated strategy use, metacognitive monitoring and control, and motivation and meta-motivation for students who might not otherwise do so spontaneously. For instance, in-game quests could be presented in a multiple-play format with increasing levels of difficulty in order to develop mastery. The autonomy of choosing

the number of gameplay sessions that evolve with increasing levels of challenge could facilitate self-efficacy and depth of knowledge. Monitoring training could be presented within the narrative of the game by having students make judgments at critical junctures in the problem-solving progress. These judgments could be presented through characters or technology devices (e.g., tablet, phone, and so on) with points awarded for accuracy. This approach would maintain the narrative by seamlessly integrating tasks within the normal sequence of gameplay events. Finally, findings related to the use of the diagnostic worksheet have shown how critical it is for students to use in-game cognitive tools in a strategic manner. The use of tools similar to the diagnostic worksheet that assist in off-loading and organizing information pertinent to successful performance in the environment can be scaffolded in such a manner (e.g., through character dialogue) so as to keep students engaged within the narrative of the game.

The second major finding in the current investigation was that Crystal Island–Outbreak supported content learning equally for both genders despite boys’ incoming advantages in gaming experience. Overall, playing Crystal Island–Outbreak led to significant content learning gains for both boys and girls. This finding of equivalent performance across genders is significant from an instructional design perspective because modifications in the environment were not necessary to compensate for pre-existing student differences. Rather, the students themselves appeared to adapt. For instance, boys entered the study with higher perceived skill and experience with computer games than girls. Accordingly, they showed an advantage over girls in the early stages of Crystal Island–Outbreak with respect to game score, but this gap had closed to the point of statistically nonsignificant differences by the end of play. Thus, aspects of the game design, characteristics of the female students themselves, or an interaction of the two compensated for this lack of experience. The game itself was not altered to accommodate gender gaming preferences (Arroyo et al., 2013) yet girls in this study overcame initial handicaps while showing equivalent levels of situational interest as boys. This begs an interesting question for future research: What factors compensate for girls’ lack of gaming experience, perceived skill, and early gameplay disadvantages? In the current study, it did not appear to be prior knowledge, self-report goal orientation, or situational interest that assisted girls. However, it is possible that girls were motivated by the academic nature of the task posed within the game setting and also by the narrative aspect of Crystal Island–Outbreak (Cassell & Jenkins, 1998; Meece & Painter, 2008).

Another possible interpretation of the gender-related findings we have described is that gender functioned as a proxy for basic skills that could be learned in a relatively short amount of time. If true, this begs the question of how interesting are gender differences that can be eradicated within 1 hr of interaction in an environment? This is a valid concern, and we would concur that the importance of such differences as “interesting” is minimal. Rather, we would argue that the *meaningfulness* of such a finding is important, given both the literature showing overwhelming gender differences in gaming experience for boys (Rideout, Foehr, & Roberts, 2010) and supporting evidence of such discrepancies from this study reflected by the self-report findings (gaming skill, experience) in this study. It is important to demonstrate that games can be built that are essentially equivalent learning platforms for

either gender and that such incoming differences can be ameliorated with appropriate instructional design.

The third major finding in this study was that while students of both genders performed at similar levels, their self-regulatory approach appeared to vary significantly. These differences were illustrated through the three primary facets of SRL (Zimmerman, 2000). As hypothesized, boys showed advantages using an in-game cognitive tool, the diagnostic worksheet, as evidenced by their higher scores on the diagnostic facet of the worksheet. It is important to note that these advantages disappeared when controlling for reported prior experience playing games. Nevertheless, it is reasonable to posit that schemas developed in prior gaming experiences provided a framework for using tools such as the diagnosis worksheet that function to track progress and also off-load information important to obtaining the overall goal (Sweller, Van Merriënboer, & Paas, 1998).

Interesting similarities and differences also emerged between genders for metacognitive monitoring. No significant differences were found between genders for either calibration or response bias. This finding replicates those by Lundeberg, Fox, Brown, and Elbedour (2000) who found similar levels of confidence and calibration for university students in a traditional academic setting. However, within-group differences in response bias for boys were particularly predictive of performance in Crystal Island–Outbreak and microbiology content learning. Boys who were underconfident performed significantly better than their overconfident peers. The implications of being overconfident and having illusionary superiority (Kruger & Dunning, 1999) have been well documented. Our finding of differences in the predictive nature of the direction of judgment bias between genders should be explored further to test the generalizability of this finding and the extent to which this effect for boys is unique to gaming environments.

Our hypothesis that situational interest would predict performance better for boys than girls was partially confirmed. Situational interest approached significance as a unique predictor of in-game performance for boys but not girls. Moreover, boys, but not girls, who solved the mystery had higher situational interest than their same-gender peers who did not solve the mystery. As a whole, findings in the current study leave us somewhat lacking in our understanding of what motivational factors influence girls’ performance in game-based learning environments. One possibility is that we did not measure the appropriate variables that would predict performance. Perhaps other measures that were not included, such as task value (Eccles & Wigfield, 1995), would reveal significant relationships.

The factors that contributed to success for girls in Crystal Island–Outbreak were more elusive not only with regard to motivation but also with regard to prior knowledge. Prior knowledge appeared to impact game performance for boys but not girls. Boys showed a significant relationship ($r = .35$) between prior knowledge and in-game score; however, no such relationships were found for girls. Exactly why this was the case is difficult to explain; however, one possibility is that boys’ advantages in computer gaming experience provided them with automaticity with controls and more developed schemas for gaming environments. Therefore, it is possible that boys utilized their prior knowledge more readily than girls because their resource demands were less with regard to adapting to the game. Games that extend the duration of gameplay over many sessions could examine if these

gender differences generalize and, if so, whether they decrease over time as girls become more familiar with the game environment.

Few relationships were found in this study between goal orientation and content-based or game-based performance. This finding is contrary to that of Winters, Greene, and Costich (2008) in their review of computer-based learning environments, who found that a mastery goal orientation led to an increased use of cognitive strategies. We found no correlation between mastery approach (or performance approach) scores and our measure of cognitive tool usage (diagnosis worksheet). This lack of relationship held when both genders were analyzed separately. Given the limitations of relying solely on self-report data (Winne, 2010), one potential avenue for research would be to examine measures of trace data that coincide with particular goal orientations. For instance, game-based learning environments might be constructed that allow students to practice various tasks or quests through multiple iterations in order to develop higher levels of mastery. This would provide the opportunity to make direct comparisons of trace and self-report measures.

Limitations

While this study contributed to our understanding of the impact of performance by gender in one game-based learning environment, there are a number of limitations that should be considered. In this study, we were able to observe SRL in its most commonly studied context, that of a single episode or task. There is greater potential for designs that include expanded sampling over time to observe changes in SRL and aptitude qualities of SRL (Jackson, & McNamara, 2013; Winne, 2010; Winne & Perry, 2000). In addition, there are inherent weaknesses and threats to validity when relying solely upon self-report measures or even a single measure to indicate levels of complex constructs (Schunk, 2008; Winne, 2010). Adding additional measures—for instance traceable measures of goal orientation—could provide greater assurances for results. Moreover, the measures of metacognition were limited to calibration and response bias judgments. Additional measures of metacognition such as the regulation of strategy use and tactics within Crystal Island–Outbreak would be informative and more representative of the construct as a whole. Also, internal reliability was not possible to compute for in-game multiple-choice items due to the manner in which the data were stored as totals across items presented from each character rather than as item-level data. With regard to our measure of content knowledge, even though feedback was not provided to students, we cannot rule out the possibility of testing effects impacting the increase in scores (Cull, 2000). Finally, given that perceived skill and frequency of game play were measured with single Likert scale items from the gaming survey, the effects of these variables may have been limited because of the paucity of items and the potential lack of reliability by single-item measures.

References

- Adams, D. M., Mayer, R. E., MacNamara, A., Koenig, A., & Wainess, R. (2012). Narrative games for learning: Testing the discovery and narrative hypotheses. *Journal of Educational Psychology, 104*, 235–249. doi:10.1037/a0025595
- Ainley, M., & Ainley, J. (2011). Student engagement with science in early adolescence: The contribution of enjoyment to students' continuing interest in learning about science. *Contemporary Educational Psychology, 36*, 4–12. doi:10.1016/j.cedpsych.2010.08.001
- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*, 261–271. doi:10.1037/0022-0663.84.3.261
- Anderman, E. M., & Young, A. J. (1994). Motivation and strategy use in science: Individual differences and classroom effects. *Journal of Research in Science Teaching, 31*, 811–831. doi:10.1002/tea.3660310805
- Arroyo, I., Bursleson, W., Tai, M., Muldner, K., & Woolf, B. P. (2013). Gender differences in the use and benefit of advanced learning technologies for mathematics. *Journal of Educational Psychology, 105*, 957–969. doi:10.1037/a0032748
- Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology, 96*, 523–535. doi:10.1037/0022-0663.96.3.523
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.
- Barab, S., Dodge, T., Thomas, M. K., Jackson, C., & Tuzun, H. (2007). Our designs and the social agendas they carry. *Journal of the Learning Sciences, 16*, 263–305. doi:10.1080/10508400701193713
- Britner, S. L., & Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching, 43*, 485–499. doi:10.1002/tea.20131
- Butler, D., & Winne, P. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*, 245–281. doi:10.3102/00346543065003245
- Carr, D., & Pelletier, C. (2009). Games, gender, and representation. In R. Ferdig (Ed.), *Handbook of research on effective electronic gaming in education* (Vol. 2, pp. 911–921). Hershey, PA: Information Science Reference/IGI Global.
- Carr, M., Jessup, D. L., & Fuller, D. (1999). Gender differences in first-grade mathematics strategy use: Parent and teacher contributions. *Journal for Research in Mathematics Education, 30*, 20–46. doi:10.2307/749628
- Cassell, J., & Jenkins, H. (1998). *From Barbie to Mortal Kombat: Gender and computer games*. Cambridge, MA: MIT Press.
- Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education, 59*, 661–686. doi:10.1016/j.compedu.2012.03.004
- Corno, L. (1993). The best-laid plans: Modern conceptions of volition and educational research. *Educational Researcher, 22*, 14–22. doi:10.3102/0013189X022002014
- Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology, 14*, 215–235. doi:10.1002/(SICI)1099-0720(200005/06)14:3<215::AID-ACP640>3.0.CO;2-1
- D'Mello, S. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology, 105*, 1082–1099. doi:10.1037/a0032674
- Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Beverly Hills, CA: Sage.
- Durndell, A., & Haag, Z. (2002). Computer self-efficacy, computer anxiety, attitudes towards the Internet, and reported experience with the Internet, by gender, in an East European sample. *Computers in Human Behavior, 18*, 521–535. doi:10.1016/S0747-5632(02)00006-7
- Dweck, C. (1986). Motivational processes affecting learning. *American Psychologist, 41*, 1040–1048. doi:10.1037/0003-066X.41.10.1040
- Eccles, J., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs.

- Personality and Social Psychology Bulletin*, 21, 215–225. doi:10.1177/0146167295213003
- Elliot, A. J., & McGregor, H. A. (2001). A 2×2 achievement goal framework. *Journal of Personality and Social Psychology*, 80, 501–519. doi:10.1037/0022-3514.80.3.501
- Elliott, E. S., & Dweck, C. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology*, 54, 5–12. doi:10.1037/0022-3514.54.1.5
- Finney, S. J., Pieper, S. L., & Barron, K. E. (2004). Examining the psychometric properties of the achievement goal questionnaire in a general academic context. *Educational and Psychological Measurement*, 64, 365–382. doi:10.1177/0013164403258465
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74, 59–109. doi:10.3102/00346543074001059
- Greenberg, B. S., Sherry, J., Lachlan, K., Lucas, K., & Holmstrom, A. (2010). Orientations to video games among gender and age groups. *Simulation & Gaming*, 41, 238–259. doi:10.1177/1046878108319930
- Hamlen, K. R. (2011). Children's choices and strategies in video games. *Computers in Human Behavior*, 27, 532–539. doi:10.1016/j.chb.2010.10.001
- Hayes, E. R. (2011). Gender and gaming. In S. Tobias & J. Fletcher (Eds.), *Computer games and instruction* (pp. 453–476). Charlotte, NC: Information Age.
- Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, 41, 111–127. doi:10.1207/s15326985Sep4102_4
- Jackson, G. T., & McNamara, D. S. (2013). Motivation and performance in a game-based intelligent tutoring system. *Journal of Educational Psychology*, 105, 1036–1049. doi:10.1037/a0032580
- Johnson, A. M., Azevedo, R., & D'Mello, S. K. (2011). The temporal and dynamic nature of self-regulatory processes during independent and externally assisted hypermedia learning. *Cognition and Instruction*, 29, 471–504. doi:10.1080/07370008.2011.610244
- Ketelhut, D., Dede, C., Clarke, J., Nelson, B., & Bowman, C. (2007). Studying situated learning in a multi-user virtual environment. In E. Baker, J. Dickieson, W. Wulfbeck, & H. O'Neil (Eds.), *Assessment of problem solving using simulations* (pp. 37–58). Mahwah, NJ: Erlbaum.
- Kinzie, M. B., & Joseph, R. D. (2008). Gender differences in game activity preference of middle school children: Implications for educational game design. *Educational Technology Research and Development*, 56, 643–663. doi:10.1007/s11423-007-9076-z
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41, 75–86. doi:10.1207/s15326985Sep4102_1
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77, 1121–1134. doi:10.1037/0022-3514.77.6.1121
- Linn, M. C., & Eylon, B. S. (2006). Science education: Integrating views of learning and instruction. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 511–544). Mahwah, NJ: Erlbaum.
- Lundeberg, M. A., Fox, P. W., Brown, A. C., & Elbedour, S. (2000). Cultural influences on confidence: Country and gender. *Journal of Educational Psychology*, 92, 152–159. doi:10.1037/0022-0663.92.1.152
- Lundeberg, M., & Mohan, L. (2009). Context matters: Gender and cross-cultural differences in confidence. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Handbook of metacognition in education* (pp. 221–239). New York, NY: Routledge/Taylor & Francis.
- Mayer, R. E. (2011). Multimedia learning and games. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 281–305). Charlotte, NC: Information Age.
- McQuiggan, S., Goth, J., Ha, E. Rowe, J., & Lester, J. (2008, June). Student note-taking in narrative-centered learning environments: Individual differences and learning outcomes. In B. Woolf, E. Aimeur, R. Nkambou, & S. Lajoie (Eds.), *Proceedings of the Ninth International Conference on Intelligent Tutoring Systems (ITS 2008)*, Montreal, ON, Canada, 510–519. New York, NY: Springer.
- Meece, J. L., & Jones, M. G. (1996). Gender differences in motivation and strategy use in science: Are girls rote learners? *Journal of Research in Science Teaching*, 33, 393–406. doi:10.1002/(SICI)1098-2736(199604)33:4<393::AID-TEA3>3.0.CO;2-N
- Meece, J. L., & Painter, J. (2008). Gender, self-regulation, and motivation. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 339–368). New York, NY: Erlbaum.
- Moreno, R., & Mayer, R. E. (2005). Role of guidance, reflection, and interactivity in an agent-based multimedia game. *Journal of Educational Psychology*, 97, 117–128. doi:10.1037/0022-0663.97.1.117
- Nietfeld, J. L., Cao, L., & Osborne, J. W. (2006). The effect of distributed monitoring exercises and feedback on performance, monitoring accuracy, and self-efficacy. *Metacognition and Learning*, 1, 159–179. doi:10.1007/s10409-006-9595-6
- Nietfeld, J. L., & Shores, L. R. (2010). Self-regulation within game-based learning environments. In L. Annetta & S. Bronack (Eds.), *Serious educational game assessment: Practical methods and models for educational games, simulations, and virtual worlds* (pp. 19–42). Rotterdam, the Netherlands: Sense.
- Nietfeld, J. L., Shores, L. R., & Hoffmann, K. F. (2011, May). Learning and gender differences in a narrative-centered learning environment. In A. Verbraeck, M. Helfert, J. Cordeiro, & B. Shishkov (Eds.), *Proceedings of the 3rd International Conference on Computer-Supported Education (CSEDU 2011)*, Noordwijkerhout, the Netherlands (pp. 140–144). Lisbon, Portugal: SciTePress–Science and Technology.
- O'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes: Evidence from the computer games literature. *The Curriculum Journal*, 16, 455–474. doi:10.1080/09585170500384529
- Pintrich, P. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). San Diego, CA: Academic Press. doi:10.1016/B978-012109890-2/50043-3
- Plant, E. A., Baylor, A. L., Doerr, C., & Rosenberg-Kima, R. (2009). Changing middle-school students' attitudes and performance regarding engineering with computer-based social models. *Computers & Education*, 53, 209–215. doi:10.1016/j.compedu.2009.01.013
- Rideout, V. J., Foehr, U. G., & Roberts, D. F. (2010). *Generation M2: Media in the lives of 8–18-year-olds*. Menlo Park, CA: Kaiser Family Foundation.
- Roedel, T. D., Schraw, G., & Plake, B. S. (1994). Validation of a measure of learning and performance goal orientations. *Educational and Psychological Measurement*, 54, 1013–1021. doi:10.1177/0013164494054004018
- Rowe, J. P., Shores, L. R., Mott, B. W., & Lester, J. C. (2010, June). Integrating learning and engagement in narrative-centered learning environments. In V. Alevin, J. Kay, & J. Mostow (Eds.), *Proceedings of the 10th International Conference on Intelligent Tutoring Systems (ITS 2010)*, Pittsburgh, PA. New York, NY: Springer.
- Schraw, G. (1994). The effect of metacognitive knowledge on local and global monitoring. *Contemporary Educational Psychology*, 19, 143–154. doi:10.1006/ceps.1994.1013
- Schraw, G. (1997). Situational interest in literary text. *Contemporary Educational Psychology*, 22, 436–456. doi:10.1006/ceps.1997.0944

- Schraw, G. (2010). Measuring self-regulation in computer-based learning environments. *Educational Psychologist, 45*, 258–266. doi:10.1080/00461520.2010.515936
- Schraw, G., & Lehman, S. (2001). Situational interest: A review of the literature and directions for future research. *Educational Psychology Review, 13*, 23–52. doi:10.1023/A:1009004801455
- Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review, 7*, 351–371. doi:10.1007/BF02212307
- Schunk, D. H. (2008). Metacognition, self-regulation, and self-regulated learning: Research recommendations. *Educational Psychology Review, 20*, 463–467. doi:10.1007/s10648-008-9086-3
- Schunk, D., & Zimmerman, B. (2003). Self-regulation and learning. Part Two: Cognitive contributions to learning, development, and instruction. In W. Reynolds & G. Miller (Eds.), *Handbook of psychology* (Vol. 2, pp. 59–78). Hoboken, NJ: Wiley. doi:10.1002/0471264385.wei0704
- Shute, V. J., & Zapata-Rivera, D. (2012). Adaptive educational systems. In P. Durlach (Ed.), *Adaptive technologies for training and education* (pp. 7–27). New York, NY: Cambridge University Press. doi:10.1017/CBO9781139049580.004
- Skinner, E. A., & Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology, 85*, 571–581. doi:10.1037/0022-0663.85.4.571
- Skinner, E. A., Furrer, C., Marchand, G., & Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *Journal of Educational Psychology, 100*, 765–781. doi:10.1037/a0012840
- Sweller, J., Van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296. doi:10.1023/A:1022193728205
- Tobias, S., & Fletcher, J. D. (2007). What research has to say about designing computer games for learning. *Educational Technology, 47*, 20–29.
- Vogel, J. J., Vogel, D. S., Cannon-Bowers, J., Bowers, C. A., Muse, K., & Wright, M. (2006). Computer gaming and interactive simulations for learning: A meta-analysis. *Journal of Educational Computing Research, 34*, 229–243. doi:10.2190/FLHV-K4WA-WPVQ-HOYM
- Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist, 45*, 267–276. doi:10.1080/00461520.2010.517150
- Winne, P. H., & Hadwin, A. E. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 227–304). Mahwah, NJ: Erlbaum.
- Winne, P. H., & Hadwin, A. (2013). nStudy: Tracing and supporting self-regulated learning in the Internet. In R. Azevedo & V. Alevan (Eds.), *International handbook of metacognition and learning technologies* (pp. 293–310). New York, NY: Springer Science.
- Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). San Diego, CA: Academic Press. doi:10.1016/B978-012109890-2/50045-7
- Winters, F. I., Greene, J. A., & Costich, C. M. (2008). Self-regulation of learning within computer-based learning environments: A critical analysis. *Educational Psychology Review, 20*, 429–444. doi:10.1007/s10648-008-9080-9
- Wirt, J., Choy, S., Rooney, P., Provasnik, S., Sen, A., & Tobin, R. (2004). *The condition of education 2004*. Washington, DC: U.S. Department of Education.
- Wolters, C. (2004). Advancing achievement goal theory: Using goal structure and goal orientation to predict students' motivation, cognition, and achievement. *Journal of Educational Psychology, 96*, 236–250. doi:10.1037/0022-0663.96.2.236
- Wolters, C., Yu, S., & Pintrich, P. R. (1996). The relation between goal orientation and students' motivational beliefs and self-regulated learning. *Learning and Individual Differences, 8*, 211–238. doi:10.1016/S1041-6080(96)90015-1
- Wouters, P., van Nimwegen, C., Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology, 105*, 249–265. doi:10.1037/a0031311
- Wu, W. H., Chiou, W. B., Kao, H. Y., Hu, C. H. A., & Huang, S. H. (2012). Re-exploring game-assisted learning research: The perspective of learning theoretical bases. *Computers & Education, 59*, 1153–1161. doi:10.1016/j.compedu.2012.05.003
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaert, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulated learning* (pp. 13–39). San Diego, CA: Academic Press. doi:10.1016/B978-012109890-2/50031-7
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice, 41*, 64–70. doi:10.1207/s15430421tip4102_2
- Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology, 82*, 51–59. doi:10.1037/0022-0663.82.1.51

Received August 31, 2013

Revision received April 28, 2014

Accepted April 29, 2014 ■