

**PROBLEM SOLVING AND GAME-BASED LEARNING:
EFFECTS OF MIDDLE GRADE STUDENTS' HYPOTHESIS
TESTING STRATEGIES ON LEARNING OUTCOMES***

**HILLER A. SPIRES
JONATHAN P. ROWE
BRADFORD W. MOTT
JAMES C. LESTER**

North Carolina State University

ABSTRACT

Targeted as a highly desired skill for contemporary work and life, problem solving is central to game-based learning research. In this study, middle grade students achieved significant learning gains from gameplay interactions that required solving a science mystery based on microbiology content. Student trace data results indicated that effective exploration and navigation of the hypothesis space within a science problem-solving task was predictive of student science content learning and in-game performance. Students who selected a higher proportion of appropriate hypotheses demonstrated greater learning gains and completed more in-game goals. Students providing correct explanations for hypothesis selection completed more in-game goals; however, providing the correct explanation for hypothesis selection did not account for greater learning gains. From the analysis, we concluded that hypothesis testing strategies play a central role in game-based learning environments that involve problem-solving tasks, thereby demonstrating strong connections to science content learning and in-game performance.

*This research was supported by the National Science Foundation under Grants REC-0632450 and DRL-0822200. 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 National Science Foundation.

In 2006, the educational research community received a mandate from the National Summit on Educational Games to ramp up empirical investigations targeting how and under what conditions games can be used to maximize learning potential. One of the assumptions that members of the Summit made was that students “acquire new knowledge and complex skills from game play, suggesting gaming could help address one of the nation’s most pressing needs—strengthening our system of education and preparing workers for 21st century jobs” (Federation of American Scientists, 2006, p. 3). Games are often touted as important tools for teaching an array of 21st century skills because: (a) they accommodate various learning styles and promote a complex decision-making context (Squire, 2006); (b) skill sets and dispositions embedded in well-designed games are a good match for the contemporary, technology-rich worlds students inhabit (Gee, 2003; Spires, 2008); and (c) games have the potential to promote the 21st century skills recognized as critical for all citizens (National Research Council, 2010). Increasingly, educators, private industry as well as policy makers agree that the bottom line for success in contemporary life and work is the ability “to learn rapidly and efficiently and to go into almost any situation and figure out what has to be learned” (Morrison, 2001, para. 22). Games are often viewed as potential tools for learning since they can simulate real-world complexity and fast-paced processing in ways that traditional school learning scenarios cannot approximate.

Historically, games have been associated with *play* outside of school contexts, but in light of low testing scores and increasing rates of high school dropouts, many are looking to games for educational benefits in schools. Pelletier (2009) has challenged educators to resist the urge to frame games as a way to salvage education by claiming that the value of games should be rethought in terms of “the situated signification of ‘game’ rather than games causing learning” (p. 83). Klopfer, Osterweil, and Salen (2009) in *Moving Games Forward: Obstacles, Opportunities, & Openness*, on the other hand, provide a conceptual path for people and organizations interested in fostering the development of games for learning purposes. Resisting the urge to make an either/or argument, they make “a case for learning games grounded in the principles of good fun and good learning” (p. 1) and devote their efforts to motivating and informing educators and researchers who want to constructively participate, as creators and consumers, in the gaming domain.

Notwithstanding the growing trend for games to be fun and educational, empirical support for how games can improve learning outcomes for academic content in schools has been slow to emerge (Hayes, 2005). Mayer and Johnson (2010) conducted a recent review where they synthesized learning outcome research in three categories: “cognitive consequences, media comparisons, and value added” (p. 246). They concluded that the value added research (i.e., when researchers ask which features add value in terms of educational effectiveness of the game) is most important when the goal is academic learning. Mayer and

Johnson's (2010) research synthesis provides a much needed (a) utilitarian analysis of game effectiveness, and (b) critical definition and articulation of learning outcomes as the field of game-based learning continues to evolve.

One of the most promising goals for games is to foster problem-solving skills. Problem solving is considered a critical 21st skill by many groups, including the Partnership for 21st Century Skills (2004), which claims that the capacity to problem-solve separates students who are prepared for increasingly complex life and work environments from those who are not. Problem solving is also supported by research conducted by Levy and Murnane (2004), who, after conducting a content analysis of emerging work skills, concluded that the nation's challenge was to prepare youth for the high-wage/high-skilled jobs that involve *expert problem-solving skills and complex communication*.

Despite the importance of problem-solving skills to contemporary society, they are challenging to teach effectively (Jonassen, 2006). In contrast to textbook treatments of problem solving as isolated investigations, games can evoke student engagement in dimensional problem solving related to scientific inquiry. For example, Chris Dede and his colleagues have used multi-user virtual environments (MUVEs) as a pedagogical apparatus to teach science concepts to middle grade students. They argue that rather than learning by listening to lectures or reading textbooks, students can learn science by exploring and solving problems in realistic environments. Their early MUVE, River City, is set in a historical town in the late 1800s, where residents are becoming ill. The students take on the role of 21st century scientists who travel back in time to help the mayor identify the cause of the illness. Ketelhut (2007) found that embedding science inquiry curricula in River City could act as a catalyst for change in middle grade students' self-efficacy and learning processes. Nelson (2007) found differences in overall learning outcomes by gender with girls taking advantage of using more in-game scaffolds for learning content. Dede's latest project, EcoMUVE, is designed as an inquiry-based, simulated ecosystem experience to support learners developing an understanding of complex causality (Metcalf, Dede, Grotzer, & Kamarainen, 2010); it is an example of engaging students in problem-solving processes within a game context that uses science content.

Similar to River City, in terms of a problem-solving focus, CRYSTAL ISLAND is an example of an academic innovation that targets science education for eighth-grade middle school students. Taking their cues from Bruner (1990, p. 35), who observed that the way people organize their experience and knowledge with the social world "is narrative rather than conceptual," the CRYSTAL ISLAND designers devised a narrative-centered learning environment for students to explore concepts related to microbiology (Mott & Lester, 2006). Barab, Gresalfi, and Ingram-Goble (2010) define narrative-centered learning games as environments that "afford dynamic interplay between player and story line, between knower and known, between action and understanding" (p. 525).

THE CASE OF CRYSTAL ISLAND

CRYSTAL ISLAND (see Figure 1) is a narrative-centered learning environment built on Valve Software's Source™ engine, the 3D game platform for Half-Life 2. The curriculum underlying CRYSTAL ISLAND's mystery narrative is aligned with North Carolina's standard course of study for eighth-grade microbiology. Students play the role of the protagonist, Alyx, who is attempting to discover the identity and source of an infectious disease plaguing a research station. Several of the team members have fallen gravely ill, and it is the student's task to discover the nature and cause of the outbreak.

CRYSTAL ISLAND's narrative takes place in a small research camp situated on a recently discovered tropical island. As students explore the camp, they investigate the island's spreading illness by forming questions, generating hypotheses, collecting data, and testing hypotheses. Throughout their investigations, students interact with non-player characters offering clues and relevant microbiology facts via multimodal "dialogues" delivered through student menu choices and characters' spoken language. The dialogues' content is supplemented with virtual books, posters, and other resources encountered in several of the camp's locations. As students gather useful information, they have access to a personal digital assistant to take and review notes, consult a microbiology field manual, communicate with characters, and report progress in solving the mystery. To solve the mystery, students complete a diagnostic worksheet to manage their working hypotheses and record findings about patients' symptoms and medical histories, as well as any findings from tests conducted in the camp's laboratory. Once a student enters a hypothesized diagnosis, cause of illness, and treatment plan into the diagnosis worksheet, the findings are submitted to the camp nurse for review and possible revision.

THEORETICAL FRAMEWORK: NARRATIVE, SOCIAL, AND COGNITIVE ASSUMPTIONS FOR LEARNING WITHIN CRYSTAL ISLAND'S GAMEPLAY

Three theories provide the foundation for the development of CRYSTAL ISLAND and the specific study at hand: narrative-centered learning, activity theory, and cognitive load theory. Following is a brief discussion of each theory and an example of how the theory is enacted in relation to the game world of CRYSTAL ISLAND—specifically the problem-solving task that learners engage in to solve the mystery. We propose a blending of the three theories to illustrate how within the problem-solving task learners simultaneously engage the narrative, navigate mediating artifacts, and use tools to offset cognitive load during processing as they test their hypotheses about what is causing people to become sick.

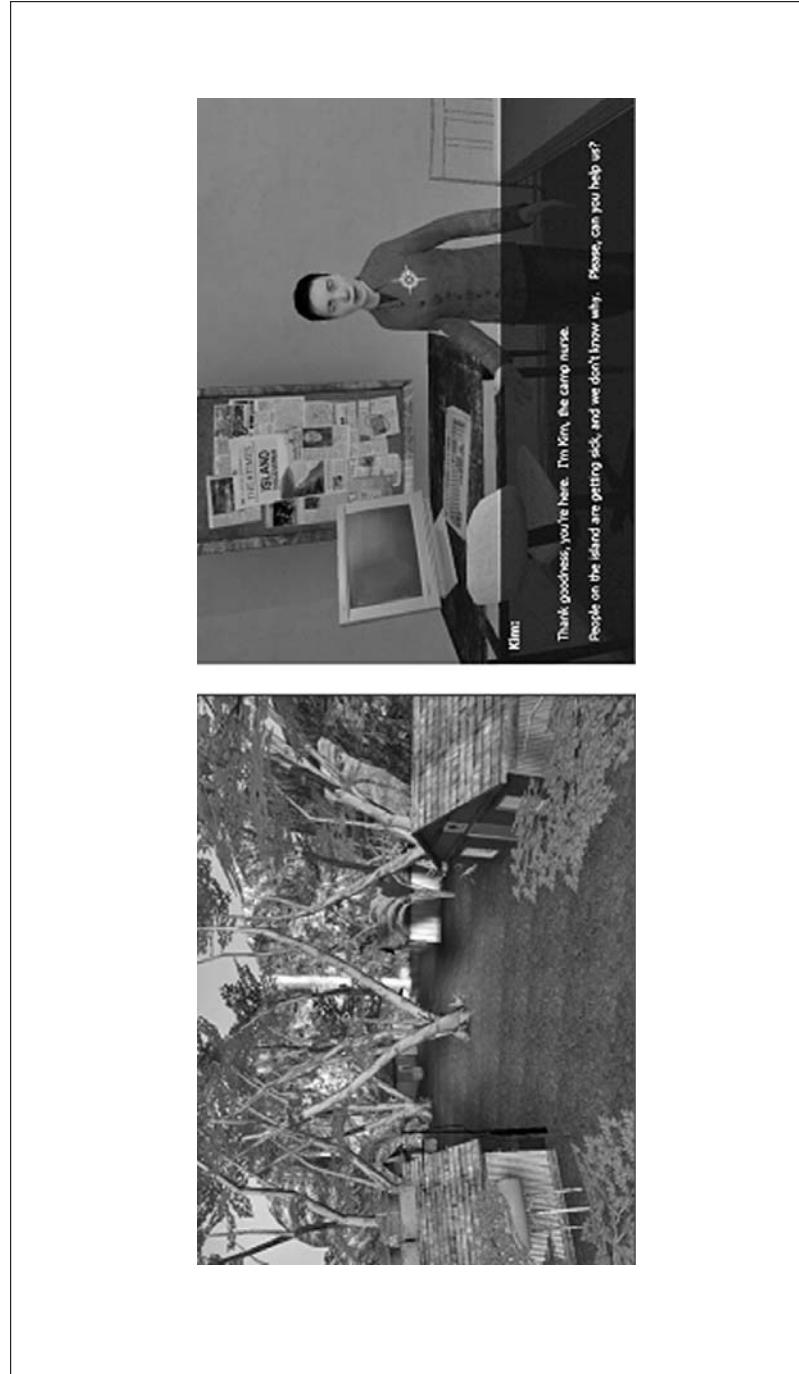


Figure 1. CRYSTAL ISLAND learning environment.

Narrative-Centered Learning

Mott, Callaway, Zettlemoyer, Lee, and Lester (1999) introduced the theory of narrative-centered learning to game-based learning environments and virtual worlds by building on Gerrig's (1993) two principles of cognitive processes in narrative comprehension. First, readers are transported; they are somehow taken to another place and time in a manner that is so compelling it seems real. Second, they perform the narrative. Simulating actors in a play, readers actively draw inferences and experience emotions prompted from interactions with the narrative text, or what Deslandes (2004) refers to as *emoting by proxy*. In the same way that good readers employ a particular stance to achieve their reading purpose and goals, a game player may employ a stance—not unlike those in the efferent-aesthetic continuum in Rosenblatt's (2004) transactional theory—in order to “read” and participate successfully in the game (Spires, Turner, Rowe, Mott, & Lester, 2010). With the recent interest in game-based learning environments and virtual world creation, narrative is being appropriated as a dynamic tool for exploring the structure and processes of game-based learning related to engagement and meaning creation (Barab et al., 2010; Mott & Lester, 2006).

Narrative-centered learning plays a central role in CRYSTAL ISLAND; the story line creates a sense of urgency by drawing the learner into the problem as explained by the character Kim, the nurse, who says, “Thank goodness, you’re here. I’m Kim, the camp nurse. People on the island are getting sick, and we don’t know why. Please, can you help us?” This interaction quickly prompts the student to a level of action. Throughout the game, Kim serves the function of increasing intensity within the narrative as she directs the learner to possible avenues to gather pertinent information. A dialogue branch listing the sub-activities necessary for completing the task includes: “You can gather clues by talking to other team members, exploring the camp, and using the laboratory’s testing equipment. Complete a diagnosis worksheet, then come talk to me.” Students are immediately thrust into the action of the game.

Activity Theory

Activity theory suggests that learning is shaped by social practice, with people and artifacts mediating the learner’s relationship with reality (Cole & Engeström, 1993; Kaptelinin & Nardi, 2006; Vygotsky, 1978). Derived from Vygotsky’s cultural-historical approach to learning and applied to distributed learning, activity theory casts learning as “expanding involvement” (social and intellectual), with other people and cultural tools. Meaning and subsequent learning are produced in the enactment of activity with other people and things, rather than being something confined to individual mental processes alone. Activity theory’s mediational triangle (Cole & Engeström, 1993; Engeström, 1987) provides a theoretical lens for gameplay, specifically the tools available to complete the activity and the relationships within the game that create the interactivity leading

to action and subsequent meaning. The basic triangle consists of the following elements: subject, object, and mediating tools—all in the service of learning outcomes. Activity theory and the mediational triangle can be used to evaluate students' problem-solving paths and specifically the architecture of how and under what conditions students test hypotheses to solve the mystery.

Activity theory is enacted in the CRYSTAL ISLAND problem space as students become involved in shared tasks with tools that potentially can produce learning, namely, in this context, microbiology content. As students progress through the gameplay, they are free to engage the resources within the game environment: they may choose to speak with other characters, explore the island, and read books and posters, or begin conducting tests in the laboratory. The activity system produced an opportunity for learning through the provision of additional resources; Vygotsky (1978) called these types of scaffolding opportunities “zones of proximal development,” which he defined as the difference between what one could do alone and what one could do with assistance. Additional pedagogical supports for problem solving in the character dialogues are provided by Elise, the lab technician character. Activity theory also comes into play when something unanticipated happens within the problem space. For example, some students take a scattershot approach to testing hypotheses about what is causing the illness rather than a strategic approach; this choice is a result of social and cultural practices related to personal learning trajectories that students bring to the problem-solving context.

Students are encouraged to use the lab's testing equipment to investigate objects for contaminants, and to speak with Elise if any contaminated objects are found. Hypothesis options include: (a) I believe this object is contaminated with Pathogens; (b) I believe this object is contaminated with Mutagens; (c) I believe this object is contaminated with Carcinogens. (See options in Figure 2.)

If the students have exhausted all five of the initial tests on the equipment, they receive a phone call from the camp nurse. The nurse offers an opportunity to earn back more tests, but requires that they answer microbiology quiz questions in order to demonstrate that they are able to make informed decisions about which tests to run in the future. Students complete the diagnosis worksheet, which scaffolds problem solving by explicitly outlining four subcomponents of the game's overall task: recorded symptoms, laboratory testing results, beliefs about various candidate diagnoses, and a final diagnosis.

Cognitive Load Theory

Building on Miller's theory (1956), which showed that short-term memory is limited in the number of elements it can handle concomitantly, Sweller (1988) proposed cognitive load theory, which treats schemata as the cognitive structures that make up an individual's knowledge base. Rather than isolated facts, the contents of long-term memory are “sophisticated structures that permit us

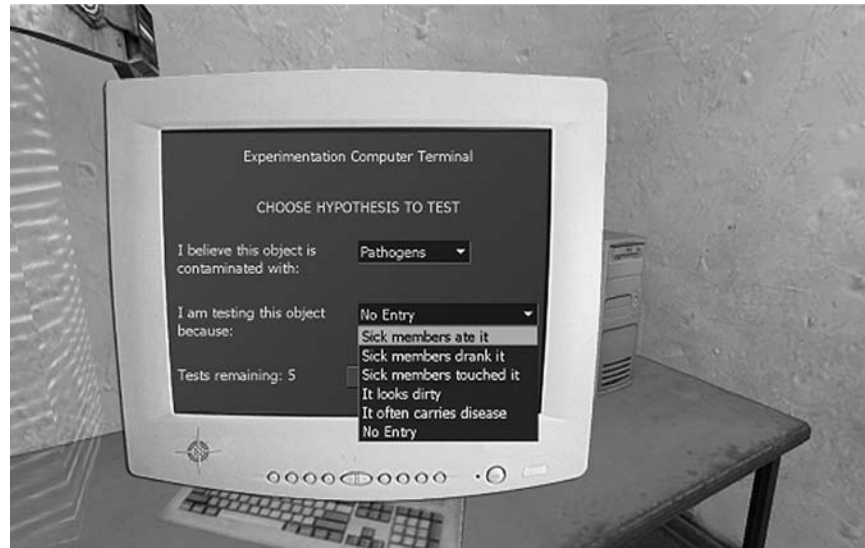


Figure 2. In-game hypothesis testing tool.

to perceive, think, and solve problems” (Sweller, 2005, p. 20). These schemata permit the learner to treat multiple elements as a single element. Mayer (2005) related cognitive load (i.e., extraneous processing) to multimedia learning in games by illustrating that extraneous processing does not enhance the desired instructional objective, and in fact may compete with cognitive processing resources.

Cognitive load theory is illustrated within *CRYSTAL ISLAND* through the strategy of having students complete the diagnosis worksheet. The worksheet serves as a tool to offset cognitive load as the students are making decisions about which objects to test as part of their hypothesis strategy. In essence, the tool was designed to help reduce working memory load and increase schema construction and automation. Employing their hypothesis testing strategies should ideally lead the students to understand the source of the team members’ sickness; for instance, students should identify that a pathogen is the cause, and more specifically, that the cause is a bacterial infection spread via the milk. Next, they must determine a specific diagnosis and treatment plan, again using the worksheet as a tool to alleviate cognitive load.

Problem solving in *CRYSTAL ISLAND* relies on students’ abilities to apply their knowledge about science and microbiology concepts (e.g., definition of

pathogen, sizes and structures of bacteria and viruses, treatments, scientific method, symptoms of various diseases). By blending three theories—narrative, activity, and cognitive load—the game design promotes engagement, understanding, motivation, and interaction through player immersion in a complex, feedback-rich problem space.

Research Context

Because problem solving is a targeted 21st century skill that is highly sought after in contemporary society (Levy & Murnane, 2004), we were curious how well students who played CRYSTAL ISLAND could problem-solve within the game. Classic models of scientific problem solving involve six steps: state the problem, form a hypothesis, test the hypothesis, collect the data, analyze the data, and draw a conclusion (Jonassen, 2006; Mayer & Wittrock, 2006). For the purposes of this investigation, we were interested in how well students could perform the combined function of forming and testing hypotheses, which is a critical feature of the CRYSTAL ISLAND gameplay. Additionally, focusing on students' hypothesis formulation and testing actions eliminated potential confounds and ensured a clear analysis with the computer trace data. Because of the level of complex processing required, we were specifically interested in the relationship between students' strategies for testing food items for contamination and their learning experiences. We focused on two facets of students' learning experiences—science content learning and in-game performance. The study addressed the following two questions: (a) What is the relationship between students' hypothesis testing strategies and their science content learning in a game environment? (b) What is the relationship between students' hypothesis testing strategies and their in-game performance?

METHOD

Participants and Design

Participants who interacted with the CRYSTAL ISLAND environment consisted of 153 eighth-grade students ranging in age from 12 to 15 ($M = 13.3$, $SD = 0.48$). Eight participants were eliminated due to incomplete data and eight were eliminated because they had prior experience with CRYSTAL ISLAND, so the final number included in the data analysis was 137 (77 males and 60 females). Approximately 3% of the participants were American Indian or Alaska Native, 2% were Asian, 32% were African American, 13% were Hispanic or Latino, and 50% were White. The study was conducted prior to students' exposure to the microbiology curriculum unit of the state's standard course of study in their regular classes. The study took place in a public magnet school, in which 35%

of the students received free or reduced lunches; the school is located in a large urban district in a southeastern state.

Measures of Learning

Multiple-Choice Content Questions

The pre- and post-intervention content test consisted of 16 questions designed by an interdisciplinary team of researchers and curriculum specialists to align with the microbiology content of the state's standard course of study. Two eighth-grade science teachers critiqued the content test to help establish content validity. There were eight factual questions that were designed to assess students' literal understanding of microbiology content and eight application questions that required students to apply microbiology knowledge to a specific situation. Questions were pilot tested on a group of students from the same population to test for content appropriateness and clarity.

In-Game Performance

A critical feature of the research was the analysis of game-based performance traces to measure students' behavioral engagement. Using automated logging facilities, students' problem-solving activities were monitored during gameplay and logged to a database for post-hoc analysis. The logging facility recorded student actions at a level of granularity that provided detailed representations of all navigation and artifact manipulation behaviors within the learning environment at millisecond intervals. This study examined one aspect of in-game performance traces, namely, the number of goals completed. To complete CRYSTAL ISLAND, participants had to complete 11 goals; however, not all students completed all of the goals in the allotted 60 minutes. Therefore, the number of goals completed per student ranged anywhere from 0 to 11.

Procedure

Each participant was seated in front of a laptop computer with headphones as they participated in a three-part introductory session that lasted 20 minutes. First, students participated in a brief demonstration session, which included general details about the CRYSTAL ISLAND mystery and game controls. One of the researchers conducted the demonstration. Second, students completed the pre-intervention content test. Third, students were provided with several CRYSTAL ISLAND supplementary documents that could be referred to during gameplay. These materials consisted of a CRYSTAL ISLAND backstory and task description, a character handout, a map of the island, and an explanation of the game's controls.

After the introductory session was completed, participants were given 60 minutes to play the game and solve the mystery. Solving the mystery consisted of several objectives, including the following: learning about pathogens, viruses,

and bacteria; compiling the symptoms and recent history of the sick researchers; recording details about diseases believed to be afflicting the team members; and formulating and testing a variety of hypotheses concerning sources for the disease. The final task was to report the solution to the mystery—which included the cause, source, and treatment—to the camp nurse. After the designated amount of time had elapsed (60 minutes) or the students had completed their interaction, participants were instructed to move on to the post-intervention phase where they completed the multiple-choice content questions. The total duration of the introduction, gameplay, and post-test sessions did not exceed 120 minutes. Students who finished before the allotted time were directed to complete a related worksheet; all students left the testing area at the same time.

RESULTS

Several analyses were conducted in order to investigate relationships among students' hypothesis testing strategies, science content learning gains, and in-game performance. First, however, we examined the extent to which pre-test scores differed from post-test scores. The results of a two-tailed, paired samples *t*-test indicated that students achieved significant learning gains from their interactions with CRYSTAL ISLAND, $t(136) = 9.34, p < .001$; the effect size ($r = .39$) is indicative of medium to large strength (Cohen, 1988). Students answered an average of 2.26 more questions correctly on the post-test ($M = 8.60, SD = 2.94$) than the pre-test ($M = 6.34, SD = 1.98$). After establishing that learning took place from pre- to post-test, multiple regression analyses were conducted to examine the relationships between hypothesis testing strategies (i.e., proportion of appropriate hypotheses and correct explanations) and two dependent variables: (a) science content learning and (b) in-game problem-solving performance.

Results for Hypothesis Testing Strategies and Science Content Learning

Prior to conducting the substantive analyses between the predictor and the outcomes, we examined the association between in-game performance (operationalized as goals completed within the game) and learning gains on the content test. Multiple regression analysis was performed in order to determine the association between number of goals completed and learning outcomes. The post-test score was used as the dependent variable and both pre-test score and the number of goals completed were treated as independent variables, thus providing estimates of unique prediction provided by both in-game performance and pre-test score. The model explained a significant proportion of the variance in post-test scores, $F(2, 134) = 33.17, p < .001$; adjusted R^2 was .32. Both pre-test score ($\beta = .32, p < .001, r = .39$) and number of goals completed ($\beta = .43, p < .001, r = .48$) demonstrated significant prediction of post-test score.

A regression analysis was conducted in order to investigate the association between students' hypothesis selection practices and learning. Students conducted tests in the laboratory using a virtual computer terminal (shown in Figure 2). Before each test, students would determine whether to test an object for pathogens, mutagens, or carcinogens, and select a short justification for the test they were about to conduct. Early in the mystery, students were expected to infer that the sick team members were suffering from some form of pathogen that had contaminated the camp's food supply. One measure of students' ability to navigate the hypothesis space in CRYSTAL ISLAND is the proportion of laboratory tests conducted that hypothesize an item as contaminated with pathogens rather than mutagens or carcinogens. Conducting laboratory tests to investigate mutagens or carcinogens on objects represents a portion of the hypothesis space that students should be able to eliminate before conducting any tests.

The proportion of appropriate hypotheses was calculated for each student by counting the number of tests conducted for pathogens, divided by the total number of tests conducted. Multiple regression analysis was performed with the post-test score as the dependent variable, and pre-test score and proportion of appropriate hypotheses as the independent variables. Together the variables explained a significant proportion of the variance in post-test scores, $R^2 = .22$, $F(2, 124) = 17.26$, $p < .001$. Both pre-test scores ($\beta = .36$, $p < .001$; $r = .39$) and appropriate hypotheses ($\beta = .25$, $p = .002$; $r = .30$) were significantly predictive of post-test scores (see Table 1). The findings indicate that students who tested a higher proportion of appropriate hypotheses exhibited increased learning gains. Further, both pre-test score and proportion of appropriate hypotheses accounted for unique variance in learning gains.

A similar analysis was conducted to investigate the number of correct explanations provided for laboratory tests. Students could justify a test by selecting one of five possible explanations: the sick team members recently ate the item, the sick team members recently drank the item, the sick team members recently touched the item, the item usually carries disease, or the item looks dirty (see

Table 1. Multiple Regression of Pre-Test Score and Proportion of Appropriate Hypotheses on Post-Test Score

Predictor	<i>B</i>	<i>SE</i>	β
Pre-test	.54**	.12	.36**
Proportion of appropriate hypotheses	2.9*	.91	.25*
Intercept	3.24	.94	

Note: R^2 Adj. = .21.

* $p < .05$.

** $p < .001$.

Figure 2). The only explanations that were consistent with the virtual characters' dialogues were the first two; the latter three were considered incorrect for the scenario. The proportion of correct explanations was calculated for each student by counting the number of tests conducted with correct explanations, divided by the total number of tests conducted. A multiple regression analysis was performed with post-test score as the dependent variable, and pre-test score and proportion of correct explanations as independent variables. While the model explained a significant proportion of the variance in post-test score, $R^2 = .17$, $F(2, 124) = 12.59$, $p < .001$, only pre-test score was found to be a significant predictor of post-test score ($\beta = .37$, $p < .001$; $r = .39$). Proportion of correct explanations was not a significant predictor of post-test score ($\beta = .12$, $p > .10$; $r = .18$).

Results for Hypothesis Testing Strategies and In-Game Performance

Analyses were conducted to investigate the association between hypothesis testing strategies and in-game performance. A multiple regression analysis was performed with number of in-game goals completed as the dependent variable, and pre-test score and proportion of appropriate hypotheses as independent variables. The model significantly predicted number of goals completed, $F(2, 124) = 9.87$, $p < .001$; adjusted R^2 was .14. Investigating each of the individual predictors, proportion of appropriate hypotheses had a significant association with number of goals completed, ($\beta = .34$, $p < .001$; $r = .35$). Pre-test score was not significantly associated with number of goals completed, ($\beta = .12$, $p > .10$; $r = .16$). Hence, proportion of appropriate hypotheses—and not pre-test scores—accounted for unique variability in the in-game performance (see Table 2).

A multiple regression analysis was performed to investigate the association between proportion of correct explanations and in-game performance. The dependent variable was number of in-game goals completed, and the independent variables were pre-test score and proportion of correct explanations. The model significantly predicted number of goals completed, $F(2, 124) = 4.28$, $p = .016$;

Table 2. Multiple Regression of Pre-Test Score and Proportion of Appropriate Hypotheses on In-Game Performance

Predictor	<i>B</i>	<i>SE</i>	β
Pre-test	.12	.08	.12
Proportion of appropriate hypotheses	2.5*	.64	.34*
Intercept	6.74	.66	

Note: R^2 Adj. = .12.

* $p < .001$.

adjusted R^2 was .05. The proportion of correct explanations had a significant association with the number of goals completed, ($\beta = .20$, $p = .027$; $r = .22$); pre-test score was not significantly associated with the number of goals completed, ($\beta = .13$, $p > .10$; $r = .16$). In other words, proportion of correct explanations—and not pre-test score—predicted variability in in-game performance.

DISCUSSION

The discussion examines to what extent the theoretical framework (i.e., narrative, activity, and cognitive load), which motivated the design of CRYSTAL ISLAND, is realized by the empirical findings in the study. Additionally, the findings are rationalized in order to characterize the relationship between students' problem-solving actions and learning experiences in the game environment. The section is organized into three categories: (a) the relationship between hypothesis testing strategies and science content learning, (b) the relationship between hypothesis testing strategies and in-game performance, and (c) limitations of the research.

Hypothesis Testing Strategies and Science Content Learning

Students who selected a higher proportion of appropriate hypotheses during gameplay exhibited greater learning gains; however, providing correct explanations for laboratory tests was not observed to have a significant association with student post-test performance. Selection of appropriate hypotheses was associated with students' ability to apply content knowledge. Specifically, hypothesis selection was associated with students' ability to distinguish between the characteristics of pathogens, mutagens, and carcinogens. If students understood the characteristics of pathogens, mutagens, and carcinogens, and inferred that the spreading illness on CRYSTAL ISLAND stemmed from a pathogen, then they should have been able to infer that they need only conduct laboratory tests for pathogen-based contaminants. This inference required the type of content knowledge that was assessed by the post-test. Students who had the ability to apply their content knowledge to effectively select hypotheses during laboratory tests had a stronger mastery of the relevant content than students who were unable to apply content knowledge to effectively select hypotheses during laboratory tests. Therefore, students who were able to effectively select hypotheses exhibited increased performance on the post-test.

Providing correct explanations was associated with students' ability to apply knowledge about the scenario. Specifically, providing explanations for laboratory tests was associated with students' ability to apply knowledge about the virtual characters' prior histories and eating habits. Students needed only to conduct tests on objects (e.g., food and drinks) from which the virtual characters recently

consumed. Characters did not report touching any objects, thus rendering weak explanations about objects looking dirty or prone to carrying diseases. Knowledge about the virtual characters' prior histories was distinct from the content knowledge that was assessed by the post-test. One would not expect to observe that students' ability to apply such knowledge would transfer to the post-test and yield increased learning gains. Therefore, it was likely that providing correct explanations would have no effect on post-test performance.

There is emerging evidence that the theories that motivated the game design may also bear on the empirical findings of the study. For example, as explained earlier, activity theory is enacted within the game and allowed for learning through the provision of additional resources (e.g., reading books and posters, and interacting with knowledgeable characters). Specifically, the in-game hypothesis testing tool (see Figure 2) served a scaffolding function for students' hypothesis formulation and testing processes. This support was designed to induce what Vygotsky (1978) called "zones of proximal development," so that students could process the microbiology content with appropriate scaffolding. In terms of our findings, students who were able to effectively use this tool demonstrated greater science content learning. In essence, the students who performed more effectively experienced "expanding involvement" with the hypothesis testing tool; this expanded involvement supported meaning making that ultimately manifested itself as science content learning.

Hypothesis Testing Strategies and In-Game Performance

Student selection of appropriate hypotheses had a positive association with their goal achievement in the game; namely, students providing correct explanations for laboratory tests demonstrated higher levels of goal achievement in the game. Students interacted with CRYSTAL ISLAND for a fixed amount of time (approximately 60 minutes). Students who selected greater proportions of appropriate hypotheses and provided greater proportions of correct explanations were more likely to be efficient at identifying the source of CRYSTAL ISLAND's spreading illness. Identifying the source of CRYSTAL ISLAND's illness required students to test a specific food object for pathogens that the sick team members recently consumed. This was the only laboratory test that would yield a positive result for contaminants. Students who practiced effective hypothesis testing strategies, and were therefore efficient at identifying the source of the spreading illness, would have more time available to complete other goals in the game (e.g., goals related to reporting their final diagnosis). Therefore, it is likely that efficient hypothesis testing strategies would be associated with an increased numbers of goals completed in the game.

As previously mentioned, some of the theories that motivated the game design align with empirical findings; in the case of in-game performance it appears there

are tentative connections with narrative and cognitive load theories. For example, narrative theory may have implications for the number of goals students' completed within game. It is plausible that students who buy into the narrative storyline of CRYSTAL ISLAND and, in essence, *perform the narrative* (Gerrig, 1993) also draw inferences and experience emotions prompted from the characters in the game. Since CRYSTAL ISLAND was designed to have students dialogue with at least one character (i.e., the nurse) before testing their hypotheses, it is possible for students to emotionally identify with the nurse; as a result of *emoting by proxy* (Deslandes, 2004) students may approach the hypothesis testing process more seriously and meaningfully. By applying the information collected to make better hypothesis testing decisions, students are more likely to solve the mystery.

In terms of cognitive load theory, extraneous processing competes with cognitive processing resources and can be a liability when a learner is acquiring new content material (Mayer, 2005). An important element in game design is creating an optimal balance between discovery elements and embedded supports to facilitate cognitive processing of the content. Problem solving in CRYSTAL ISLAND was contingent upon students' ability to apply their knowledge about science and microbiology concepts (e.g., definition of pathogen, sizes and structures of bacteria and viruses, treatments, scientific method, symptoms of various diseases). Various tools embedded within the game were designed to offset cognitive load, as the students made decisions about which objects to test as part of their hypothesis strategy. It is plausible that students who performed better within the game took advantage of these resources.

Limitations of the Research

This study has several limitations. First, there are limitations to using multiple-choice response formats to measure complex cognition and inquiry processes within an environment like CRYSTAL ISLAND. As Schaffer, Hatfield, Svarovsky, Nash, Nulty, Bagley, et al. (2009) asserted, "Assessments of digital learning need to focus on performance in context rather than on tests of abstracted and isolated skills and knowledge" (p. 34). The capacity to use trace data for analysis provides future opportunities to use evidence-centered design, in which there is alignment between learning theory and assessment method. Additionally, in future CRYSTAL ISLAND studies the use of transfer measures will assess how well students can apply the information learned in the game to a new context. Second, while the game is a narrative-centered learning environment, and key features of narrative are embedded in the pedagogical apparatus, the gameplay of CRYSTAL ISLAND does not approximate the action and visual engagement offered by high-end commercial games. Students are generally accustomed to action and visual stimulation when playing games and can be disappointed when academic games do not mimic the level of engagement and entertainment for

which they are accustomed. Third, the study was conducted as an experimental intervention that, by and large, took place outside of the instructional context of the classroom. The curriculum content was aligned with the state's standard course of study for eighth-grade science and was vetted by science teachers who are part of the project; however, due to logistical issues, students were not able to play the game at the appropriate juncture as a supplement to enhance their classroom instruction that focused on microbiology. The timing of the study potentially impacted the students' motivation to engage with the academic content because there was no apparent connection between the game and school requirements at the time of intervention.

CONCLUSION

Results indicated that the effective exploration and navigation of the hypothesis space in a problem-solving task was predictive of student learning. Specifically, students who selected a higher proportion of appropriate hypotheses demonstrated greater learning gains on the post-test and completed more in-game goals. Students who provided correct explanations for hypothesis selection completed more in-game goals; however, providing the correct explanation for hypothesis selection did not account for greater learning gains.

Results indicated that hypothesis testing strategies play a central role in narrative-centered learning environments, thus demonstrating their connections to learning gains and problem solving. These results are a significant contribution in that they align with a growing body of research that explores the educational benefits, both practical and theoretical, of problem-solving tasks within games in and out of school settings. Given the important role that problem solving plays as a 21st century skill, combined with the challenges schools encounter as they implement problem-solving curricula, CRYSTAL ISLAND offers one example of how to engage students in narrative-centered learning that also takes into account social aspects of learning and cognitive load concerns.

Future research with CRYSTAL ISLAND will focus on more in-depth analyses of in-game performance and its relationship to learning outcomes as well as different pedagogical game features. Plans are underway to create the next iteration of CRYSTAL ISLAND, including a web-based interface and additional modules related to eighth-grade science curriculum.

No single educational approach, including game-based learning, is effective across all subjects and for all students. As members of the 2006 National Summit on Educational Games suggested, game-based learning research needs to continue to focus on what works with whom and in which context. When the research community adequately addresses this concern, games will become more

compatible with school learning contexts and potentially have a greater impact on the development of students' 21st century skills.

REFERENCES

- Barab, S. A., Gresalfi, M., & Ingram-Goble, A. (2010). Transformational play: Using games to position person, content, and context. *Educational Researcher*, 39(7), 525-536.
- Bruner, J. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.
- Cole, M., & Engestrom, Y. (1993). A cultural-historical approach to distributed cognition. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations*. New York: Cambridge University Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Deslandes, J. (2004). A philosophy of emoting. *Journal of Narrative Theory*, 34(3), 335-372. doi: 10.1353/jnt.2005.0003
- Engeström, Y. (1987). *Learning by expanding: An activity theoretical approach to developmental research*. Helsinki: Orienta-Konsultit Oy.
- Federation of American Scientists. (2006). *Harnessing the power of video game for learning*. Retrieved from <http://fas.org/gamesummit/>
- Gee, J. P. (2003). *What video games can teach us about literacy and learning*. New York: Palgrave-Macmillan. doi: 10.1145/950566.950595
- Gerrig, R. (1993). *Experiencing narrative worlds: On the psychological activities of reading*. New Haven, CT: Yale University Press.
- Hayes, R. T. (2005). *The effectiveness of instructional games: A literature review and discussion*. Naval Air Warfare Center Training Systems Division (No. 2005-004), Orlando, FL.
- Jonassen, D. H. (Ed.). (2006). *Learning to solve complex, scientific problems*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Kaptelinin, V., & Nardi, B. (2006). *Acting with technology: Activity theory and interaction design*. Cambridge, MA: MIT Press.
- Ketelhut, D. J. (2007). The impact of student self-efficacy on scientific inquiry skills: An exploratory investigation in River City, a multi-user virtual environment. *The Journal of Science Education & Technology* (Feb. 2007), 99-111. doi: 10.1007/s10956-006-9038-y
- Klopfer, E., Osterweil, S., & Salen, K. (2009). *Moving learning games forward: Obstacles, opportunities, & openness*. The Education Arcade: MIT.
- Levy, F., & Murnane, R. (2004). *The new division of labor: How computers are creating the next job market*. Princeton, NJ: Princeton University Press.
- Mayer, R. E., & Johnson, C. (2010). Adding instructional features that promote learning in a game-like environment. *Journal of Educational Computing Research*, 42(3), 241-265. doi: 10.2190/EC.42.3.a
- Mayer, R. E., & Wittrock, M. C. (2006). Problem solving. In P. Alexander, P. Winne, & G. Phye (Eds.), *Handbook of educational psychology* (pp. 287-303). Mahwah, NJ: Erlbaum.
- Mayer, R. E. (Ed.). (2005). *The Cambridge handbook of multimedia learning*. New York: Cambridge University Press.

- Metcalf, S., Dede, C., Grotzer, T., & Kamarainen, A. (2010). *EcoMUVE: Design of virtual environments to address science learning goals*. Paper presented at the American Educational Research Association (AERA) Conference, Denver, CO.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81-97. doi: 10.1037/h0043158
- Morrison, T. (2001). *The work of nations—Interview with Robert Reich*. Aurora Online. Retrieved from <http://aurora.icaap.org/archive/reich.html>
- Mott, B., Callaway, C., Zettlemoyer, L., Lee, S., & Lester, J. (1999). Towards narrative-centered learning environments. In working notes of the *AAAI Fall Symposium on Narrative Intelligence*, pp. 78-82. Cape Cod, MA.
- Mott, B., & Lester, J. (2006). U-Director: A decision-theoretic narrative planning architecture for storytelling environments. In *Proceedings of the Fifth International Conference on Autonomous Agents and Multiagent Systems* (pp. 977-984). Hakodate, Japan.
- Mott, B., & Lester, J. (2006). Narrative-centered tutorial planning for inquiry-based learning environments. In *Proceedings of the Eighth International Conference on Intelligent Tutoring Systems* (pp. 675-684). Jhongli, Taiwan.
- National Research Council. (2010). *Exploring the intersection of science education and 21st century skills: A workshop summary*. M. Hilton (Ed.). Washington, DC: National Academy Press.
- Nelson, B. (2007). Exploring the use of individualized, reflective guidance in an educational multi-user virtual environment. *The Journal of Science Education and Technology*, 16(1), 83-97. doi: 10.1007/s10956-006-9039-x
- Partnership for 21st Century Skills. (2004). [Web site]. Retrieved from <http://www.21stcenturyskills.org>
- Pelletier, C. (2009). Games and learning: What's the connection? *International Journal of Learning & Media*, 1(1), 83-97. doi: 10.1162/ijlm.2009.0006
- Rosenblatt, L. M. (2004). The transactional model of reading and writing. In R. B. Ruddell, M. R. Ruddell, & H. Singer (Eds.), *Theoretical models and processes of reading* (pp. 1363-1398). Newark, DE: IRA.
- Schaffer, D. W., Hatfield, D., Svarovsky, G., Nash, P., Nulty, A., Bagley, E., et al. (2009). Epistemic network analysis: A prototype for 21st century assessment of learning. *International Journal of Learning and Media*, 1(2), 33-53. doi: 10.1162/ijlm.2009.0013
- Spires, H., Turner, K., Rowe, J., Mott, B., & Lester, J. (2010, April/May). *Effects of game-based performance on science learning: A transactional theoretical perspective*. Paper presented at AERA, Denver, CO.
- Spires, H. A. (2008). 21st century skills and serious games: Preparing the N generation. In L. A. Annetta (Ed.), *Serious educational games* (pp. 13-23). Rotterdam, The Netherlands: Sense Publishing.
- Squire, K. D. (2006). From content to context: Videogames as designed experience. *Educational Researcher*, 35(8), 19-29. doi: 10.3102/0013189x035008019
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285. doi: 10.1016/0364-0213(88)90023-7
- Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19-30). New York: Cambridge University Press.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Direct reprint requests to:

Hiller A. Spires
Friday Institute for Educational Innovation
North Carolina State University
Raleigh, NC 27695
e-mail: hiller_spires@ncsu.edu