Computers & Education 147 (2020) 103781

Contents lists available at ScienceDirect

Computers & Education

journal homepage: http://www.elsevier.com/locate/compedu

The agency effect: The impact of student agency on learning, emotions, and problem-solving behaviors in a game-based learning environment☆

Michelle Taub[a,*], Robert Sawyer[b], Andy Smith[b], Jonathan Rowe[b], Roger Azevedo[a], James Lester[b]

[a] Department of Learning Sciences and Educational Research, University of Central Florida, USA
[b] Department of Computer Science, North Carolina State University, USA

ARTICLE INFO

Keywords:
Agency in learning
Game-based learning
Learner-centered emotions
Self-regulated learning

ABSTRACT

Game-based learning environments are designed to foster high levels of student engagement and motivation during learning of complex topics. Game-based learning environments allow students freedom to navigate a space to interact with game elements that foster learning, i.e., agency. Agency has been studied in learning, and it has been demonstrated that increased student agency results in greater learning outcomes. However, it is unclear what is the level of agency that is required to demonstrate this effect, and whether this effect applies only to learning or to problem solving and affect during game-based learning as well. To investigate how the level of student agency impacts learning, problem solving, and affect, a study was conducted with 138 college students interacting with a game-based learning environment for microbiology, Crystal Island. This study is an extension of a previous study that examined the impact of agency on learning and problem-solving behaviors during game-based learning with Crystal Island. Students were randomly assigned to either a *High Agency* condition, a *Low Agency* condition, or a *No Agency* condition. It was found that students in the *Low Agency* condition achieved significantly higher normalized learning gain scores than students in the *No Agency* condition, and marginally higher normalized learning gains than the *High Agency* condition. Post-surveys of interest and presence indicated that students in the *No Agency* condition were less interested, and perceived themselves as less present in the virtual environment, than students in the other conditions. Students in the *No Agency* condition also experienced less frustration, confusion, and joy than the other agency conditions, indicating a less cognitively stimulating experience. Overall the results indicate that a moderate degree of agency provided to students in game-based learning environments leads to better learning outcomes without sacrificing interest and without yielding a negative emotional experience, demonstrating how even low levels of agency can positively impact learning, problem solving, and affect during game-based learning.

☆ *An earlier version of this study was presented at the 18th International Conference on Artificial Intelligence in Education (AIED 2017) in Wuhan, China and published in E. André, R. Baker, X. Hu, M. M. T. Rodrigo, & B. du Boulay (Eds.), *Proceedings of the* 18th *International Conference on Artificial Intelligence in Education (AIED 2017)* (pp. 335–346). Amsterdam, The Netherlands: Springer.

* Corresponding author. University of Central Florida, Department of Learning Sciences and Educational Research, College of Community Innovation and Education, PO Box 161250, Orlando, FL 32816-1250, USA.

E-mail address: michelle.taub@ucf.edu (M. Taub).

https://doi.org/10.1016/j.compedu.2019.103781
Received 25 November 2019; Accepted 3 December 2019
Available online 5 December 2019

1. Introduction

The current study is an extension of a previous study (Sawyer, Smith, Rowe, Azevedo, & Lester, 2017) that examined the relationship between student agency, learning, and problem-solving behavior in a game-based learning environment (GBLE). Specifically, we investigated how different versions of a GBLE that afford varying levels of student agency—High Agency, Low Agency, or No Agency—impact student learning outcomes. We also examined what problem-solving behaviors account for observed differences in learning between conditions, as well as the impact of the agency manipulation on cumulative counts and durations of problem-solving behaviors within the game. The current study builds upon the previous work (Sawyer et al., 2017 by expanding the sample size from 105 to 138 participants. Student problem solving is operationalized by computing the rates of key gameplay behaviors that are instrumental to in-game problem solving, such as conversational interactions with virtual characters, running tests in a virtual laboratory, reading in-game books and articles, recording important findings, and submitting candidate solutions to solve the problem scenario.

The study also extends the previous investigation by examining the relationship between student agency and engagement with the GBLE. Engagement is operationalized using three complementary measures. Automated facial expression analysis software is used to detect salient expressions of three key learning emotions during gameplay: confusion, frustration, and joy. We also investigate the relationship between student agency and reported presence and interest in the game. Student agency is frequently cited as an important motivational factor in the design of GBLEs (Plass, Homer, & Kinzer, 2015). Therefore, manipulations of student agency have strong potential to impact student affect and engagement during game-based learning.

Game-based learning is an approach to learning that aims to foster and maintain high levels of motivation and engagement (Clark, Tanner-Smith, & Killingsworth, 2016; Mayer, 2014; Shute, Rahimi, & Lu, 2019). Games have been developed to foster effective self-regulated learning (SRL) (Nietfeld, 2018; Taub, Azevedo, Bradbury, & Mudrick, in press), as research demonstrates that students often have difficulties deploying effective cognitive, affective, metacognitive, and motivational self-regulatory processes during learning (Azevedo, Taub, & Mudrick, 2018; Winne & Azevedo, 2014). The use of SRL processes can contribute to a deeper learning experience because students are playing an active role in their learning (Winne, 2018) and ensuring they understand the material.

A broad range of game-based learning environments (GBLEs) have been developed and investigated across different educational subjects, including games for math education, science education, and civics education (Easterday, Aleven, Scheines, & Carver, 2016; Kim & Ke, 2017; Ventura, Shute, & Kim, 2013; Wouters, van Nimwegen, van Oostendorp, & van der Spek, 2013). GBLEs afford opportunities to integrate rich problem-solving scenarios associated with inquiry-based learning (Sao Pedro, Baker, Gobert, Montalvo, & Nakama, 2013) with believable virtual worlds enabled by commercial game engines. These attributes are important to games' capacity to enhance student motivation and engagement as well as to provide adaptive support for improving student learning outcomes (Plass et al., 2015).

Over the past decade, there has been growing evidence that GBLEs can serve as an effective medium for learning, but the relationship between game design and student learning is complex (Clark et al., 2016; Mayer, 2014; Wouters et al., 2013). A meta-analysis by Wouters et al. (2013) found that GBLEs were more effective than conventional instructional methods, such as lectures, reading, and drill and practice, in terms of learning and retention, but they did not increase student motivation. Other meta-analyses found that games can be effective for learning, but the results applied to specific topics and were stronger for some populations of learners than others (Clark et al., 2016; Mayer, 2014). Counterexamples about the benefits of GBLEs are also available. One study showed that an award-winning educational game with popular gamification features (e.g., performance-based rewards) was less effective at promoting transferrable knowledge gains than a comparable intelligent tutoring system for algebra education (Long & Aleven, 2014). These findings raise key questions about how to most effectively design GBLEs to support student learning and engagement.

1.1. Student agency

Sawyer et al. (2017) explained that a key feature of GBLEs is their support for *student agency*. The game design literature describes agency in terms of the degree of freedom and control that is afforded to a player to perform meaningful actions in a virtual environment (Wardrip-Fruin, Mateas, Dow, & Sali, 2009). Student agency is closely related to human agency, which is characterized by one's intentionality, forethought, self-reactiveness, and self-reflectiveness, which enable a human to ensure an activity occurs (Bandura, 2001). Human agency is also related to constructs such as control (Malone & Lepper, 1987), self-determination (Ryan & Deci, 2000), and self-regulated learning (Winne, 2018; Winne & Hadwin, 1998, 2008).

As mentioned by Sawyer et al. (2017), key to student agency in GBLEs is the *perception* of freedom and control within the game environment. For example, preventing a student from exploring a virtual location or interacting with a virtual object is a constraint on student agency. Similarly, if actions in the game are not perceived as being meaningful, or the actions have only a superficial effect, then agency is reduced. However, the inverse can also be true. If a GBLE is designed in such a way that it leads the student to believe they can perform some action in the virtual environment, such as using a virtual object, even if they cannot do so in actuality, the student may still perceive themselves as having a high degree of student agency. Studies have suggested, though not uniformly, that increased student agency is associated with higher levels of involvement and improved learning outcomes in GBLEs (Sawyer et al., 2017; Rowe, Shores, Mott, & Lester, 2011; Snow, Allen, Jacovina, & McNamara, 2015). However, there are less desirable behaviors associated with high levels of student agency, such as failure to properly monitor and regulate cognitive and metacognitive processes necessary for successful learning (Winne & Azevedo, 2014).

Sawyer et al. (2017) previously discussed the consequences of too much freedom and control for students, which is often cited as a critique of discovery learning (Kirshner, Sweller, & Clark, 2006; Mayer, 2004), because providing freedom without accompanying levels of support can lead to struggles in selecting, organizing, and integrating relevant information. It has been found that as learning environments introduce more freedom and openness, there is an increased risk of irrelevant but engaging features, referred to as "seductive details", which may detract from student learning (Harp & Mayer, 1998) and encourage off-task behavior that is associated with both poorer learning outcomes and negative affective states (Baker et al., 2011; Rowe, McQuiggan, Robison, & Lester, 2009). Thus, it is important to ensure that when incorporating game mechanisms associated with higher levels of freedom and control, such as allowing students to explore an open world environment, these mechanics also align with the primary learning objectives of the activity. This is important for reducing issues related to students not engaging with crucial elements of the content (Mayer, 2004) or becoming overwhelmed by levels of autonomy that are incompatible with their ability to properly plan, monitor, and react (Winne & Hadwin, 2008). The objective of this article is to extend the research on the design of GBLEs that balance between student agency and learning by investigating how different levels of student agency impact student learning, problem solving, emotion, and reported engagement.

2. Related work

For this study, we augmented the previous study by Sawyer et al. (2017) by investigating game-based learning behaviors from a self-regulatory perspective (Azevedo et al., 2018; Taub et al., in press) to explore how agency influences student self-regulation. SRL is multi-componential in nature where these behaviors can be classified as cognitive, metacognitive, affective, or motivational. Agency has been found to impact cognitive and metacognitive processes (Metcalfe, Eich, & Miele, 2013; Snow et al., 2015), but less research has investigated the impact of agency on SRL *and* affective processes (including engagement and emotions).

2.1. Theoretical frameworks

Plass et al. (2015) propose a theory of game-based learning that focuses on the complexities of game design, and suggest the theoretical framework that should be used for the learning component be based on the specific learning process being studied. Thus, we focus on self-regulated learning as our learning construct and we focus on Winne and Hadwin's (1998, 2008; Winne, 2018) Information Processing Theory of SRL as our theoretical framework. According to this model, learning occurs throughout a series of four interdependent phases: task understanding, setting goals and planning, engaging in learning strategies, and making adaptations. Students can engage in different self-regulatory processes during each phase, where phases are cyclical in nature and can occur simultaneously (e.g., making adaptations and setting new goals). Additionally, different conditions, operations, products, evaluations, and standards can impact how students engage in self-regulatory processes during these learning phases, such as different agency conditions.

We also focus on D'Mello and Graesser's (2012a) Model of Affective Dynamics because we examined the impact of agency conditions on learner-centered emotions. According to this model, when students are engaging in a task and they meet an impasse, a state of confusion arises, which can be resolved by engaging in effective problem-solving strategies, thus returning the student to a state of task engagement. If the confusion is not resolved, however, this can lead to a state of frustration, and ultimately boredom, which is defined as complete disengagement from the task (D'Mello & Graesser, 2012a).

2.2. Agency in game-based learning

Providing learners with high levels of student agency is a deliberate design feature of many GBLEs, including previous versions of Crystal Island (Rowe et al., 2011) as well as other GBLEs, such as Quest Atlantis (Barab, Thomas, Dodge, Carteaux, & Tuzun, 2005) and Virtual Performance Assessments (Baker, Clarke-Midura, & Ocumpaugh, 2016). A prior study conducted with middle school students using an earlier version of Crystal Island found that student learning gains and in-game problem-solving performance were correlated with several components of engagement with the game, including presence and perceived interest (Rowe et al., 2011). Engagement is a construct that can be defined from three different perspectives (Fredricks, Blumenfeld, & Paris, 2004). Behavioral engagement relates to the actions a student makes during a task. Affective engagement relates to the affective response or influence within a task. Cognitive engagement relates to learning and self-regulatory involvement within a task. This distinction has also been made for developing frameworks for describing learning with multimedia (Domagk, Schwartz, & Plass, 2010) and games (Plass et al., 2015). Rowe et al. (2011) measured student engagement with Crystal Island through several pre- and post-game questionnaires as well as an in-game measure of student engagement relating student behaviors to problem-solving performance designed by experts of the GBLE. Their study found that student engagement, as measured by questionnaires and in-game behaviors, was associated with improved learning outcomes and in-game problem solving. Quest Atlantis immerses students in a 3D multiuser environment that allows student agency in an interactive narrative with educational quests for engaging in various curricular activities and learning about social issues. Studies with students aged 9–12 playing Quest Atlantis have shown significant learning over time in science and social studies (Barab, Dodge, Jackson, & Arici, 2003). Virtual Performance Assessments consists of an immersive 3D environment that presents middle school students with a science inquiry scenario, such as determining the cause of a mutation among a population of frogs and has been previously validated to ensure it assesses performance of science inquiry (Scalise & Clarke-Midura, 2014). These GBLEs allow students to freely explore an open virtual world while solving complex science problem scenarios; they provide examples of high student agency, but they do not address how varying levels of student agency impact the learning and motivational outcomes of students.

While high student agency has been a focus for creating immersive and engaging GBLEs, recent work, including the earlier version of this study (Sawyer et al., 2017), has begun to explore agency as a game design feature, which can be manipulated in controlled experimental settings. For example, Snow et al. (2015) investigated student agency in iSTART-2, an interactive tutor with game-like features, by analyzing college students' choice patterns in the environment. Results showed that student success was closely related to their ability to exercise controlled choice patterns as opposed to disorganized (i.e., random) choice patterns (Snow et al., 2015). More specifically, this work found that students who exhibited controlled interactions when given full agency produced the highest quality self-explanations, which suggests agency has important implications for student performance within adaptive environments. Metcalfe et al. (2013) found that proximal actions appear to affect judgments of agency to a greater extent and in a more direct way than distal variables associated with the consequences of actions. Specifically, they found that reducing participants' direct control over actions by introducing a noisy control system, where the system may or may not actually perform the action the participant is trying to perform, impacted judgments of agency more negatively than reducing the impact of those actions by introducing a noisy reward system, where participants may or may not receive a reward for a successful action. Calvert et al. found that preschool children were more attentive to computer-presented content when they had control over the content compared to sharing control with an adult or having an adult control the content completely (Calvert, Strong, & Gallagher, 2005). Overall, their results suggest user control is a central engagement feature that allows extended attention and greater interest in computer activities among young children. A recent study conducted a controlled experiment manipulating student agency in a game for learning mathematics for fifth and sixth graders, Decimal Point, and found that both agency conditions achieved similar learning gains and enjoyment (Nguyen, Harpstead, Wang, & McLaren, 2018). Students selected the order in which they performed educational mini-games, and students in the high agency condition selected similar paths to those in the low agency condition, which possibly explained the lack of difference.

Although these works explore the effects of agency on game-based learning, they lack immersive environments in which seductive details may detract from learning (Harp & Mayer, 1998) and lead to off-task behaviors, leading to lower learning outcomes (Rowe et al., 2009). The current study addresses this gap in the literature by exploring the impacts of agency on an immersive GBLE where a reduced agency version may promote learning outcomes, potentially at the cost of student engagement and motivation. The previous version of this study investigated the impact of level of agency on student learning and problem-solving behaviors (Sawyer et al., 2017). This study also investigates the impact of agency on student emotions and self-reported presence and interest.

While there is limited work investigating the impacts of agency in immersive GBLEs, notable work conducted by Veinott et al. (2013) explored this topic. Their work suggests that there may not be a significant benefit of cognitive engagement through increased student agency. They found that participants who watched an instructional video were comparably engaged as participants who played an immersive video game, Heuristica, to train decision making in the face of cognitive biases. Students in the video condition did not get to play the video game, which serves a similar purpose as the *No Agency* condition in this study (see below). However, the video used by Veinott et al. cannot be viewed as a no-agency version of the game since it was not based on the game itself but on similar material, so there is not a direct comparison of a video of someone else playing the game, which would more closely resemble a condition in which the participant had no agency. The work presented here devises a no agency condition, which is based on the same GBLE from other conditions by presenting an expert playthrough, which provides a comparable example of no student agency to the other agency conditions.

2.3. Affect in game-based learning

In addition to the important role that agency can play during learning with GBLEs, as demonstrated by Sawyer et al. (2017), affect has been found to play a large role during game-based learning as well. While agency relates to the control learners have in making behavioral decisions, affect (e.g., emotions, engagement) can impact how these decisions are made. Affect has been found to play a significant role on learning, self-regulation, and motivational outcomes of interactions with learning technologies (D'Mello & Graesser, 2012a; Ekman, 1984). Therefore, investigating the impact of agency on student affect becomes an important consideration in comparing versions of a GBLE differing in agency. For example, students who effectively resolve their confusion by engaging in effortful cognitive processing and problem solving show higher learning outcomes than if they had not been confused at all (D'Mello, Lehman, Pekrun, & Graesser, 2014). Studies have also found that frustration is a commonly occurring negative emotion during learning and can lead to disengagement with the learning material (Baker, D'Mello, Rodrigo, & Graesser, 2010). Conversely, when students accomplish goals and resolve challenges they tend to experience positive emotions such as joy (D'Mello & Graesser, 2012a). Recently, facial expression recognition has been used both in student modeling of cognitive performance and for driving interventions to promote learning outcomes in adaptive learning technologies (Calvo, D'Mello, Gratch, & Kappas, 2015). While previous work has shown that GBLEs can effectively promote positive affect (Sabourin & Lester, 2014) and use student affect to drive interventions that promote learning (D'Mello & Graesser, 2012b), there has been limited research conducted regarding the impact of agency on student affect during the learning process (Nguyen et al., 2018).

3. Study design

Overall, this study complements the previous study conducted by Sawyer et al. (2017) by providing a more general investigation of the impact of student agency on students' cognitive, affective, metacognitive, and motivational processes during game-based learning. Findings points toward design implications for the creation of GBLEs that are effective at supporting student learning and engagement. To that end, the current study investigated the following three research questions:

Research Question 1: How is learning affected by different agency conditions?

Research Question 2: How are the problem-solving behaviors related to scientific reasoning of students affected by different agency conditions?

Research Question 3: How is the user experience in terms of students' emotions, presence, and interest impacted by different agency conditions?

3.1. A game-based learning environment for fostering SRL and microbiology learning

The current study used Crystal Island, a GBLE for microbiology and literacy education developed at North Carolina State University, which has been examined in multiple research studies (Sawyer et al., 2017; Mott & Lester, 2006; Rowe et al., 2011; Taub, Azevedo, Bradbury, Millar, & Lester, 2018). The GBLE has been utilized by more than 4000 students over the past decade in studies conducted in a range of laboratory and classroom settings. It was developed with the Unity game engine and integrates science inquiry and literacy learning with a strong emphasis on reading complex informational texts about microbiology.

Crystal Island's story takes place at a remote research station on a small, tropical island. In the game, students adopt the role of a medical detective who is tasked with investigating an unidentified outbreak that has spread among a team of scientists at the virtual research station. Students explore the island from a first-person perspective, gathering information to devise an evidence-based diagnosis of the identity and transmission source of the disease. The investigation takes place across several different locations, including a Tutorial area, Infirmary, Living Quarters, Laboratory, Dining Hall, and Lead Scientist's Residence. Throughout these locations, students engage in conversational interactions with a cast of virtual characters, including sick scientists who describe their symptoms and recent medical history. They read virtual books, posters, and articles about relevant microbiology concepts, such as pathogens, viral disease, bacteria, and how disease spreads. Immediately afterward, students complete embedded assessments of reading comprehension, and then apply the knowledge they have gained toward diagnosing the spreading disease. Students gather data about the disease's transmission source by conducting tests in the virtual laboratory. As they gather information, students record their findings in an in-game diagnosis worksheet that serves as a graphical organizer for key evidence that is relevant to the problem scenario. To solve the mystery, students must specify the correct disease, transmission source, and treatment/prevention plan for the outbreak. Once students have completed their diagnosis worksheet, they attempt to solve the mystery by submitting their conclusions to a virtual nurse who resides in the island's infirmary. The nurse provides feedback about the students' diagnosis, or if it is correct, congratulates the student on solving the mystery.

3.2. Experimental conditions

To investigate the impact of student agency on their learning processes, multiple versions of Crystal Island were developed. Three alternative forms of student interaction with the Crystal Island game-based learning environment were examined in a prior study (Sawyer et al., 2017), as well as in this study: a *High Agency* condition, a *Low Agency* condition, and a *No Agency* condition. In the *High Agency* condition, the game allowed students to move freely throughout the virtual environment after they had completed a brief gameplay tutorial near the entrance to the island. Students had total control to navigate between buildings; explore the island's exterior locations; engage virtual characters in conversation; manipulate virtual objects; examine books, articles, and posters throughout the island; conduct tests in the virtual laboratory; and attempt to solve the mystery by submitting a diagnosis to the camp nurse. The freedom of movement afforded in the *High Agency* condition is depicted in Fig. 1.

In the *Low Agency* condition, students interacted with a modified version of Crystal Island that introduced several constraints on the range of possible problem-solving paths that students could take in the game. The narrative, problem scenario, virtual environment,

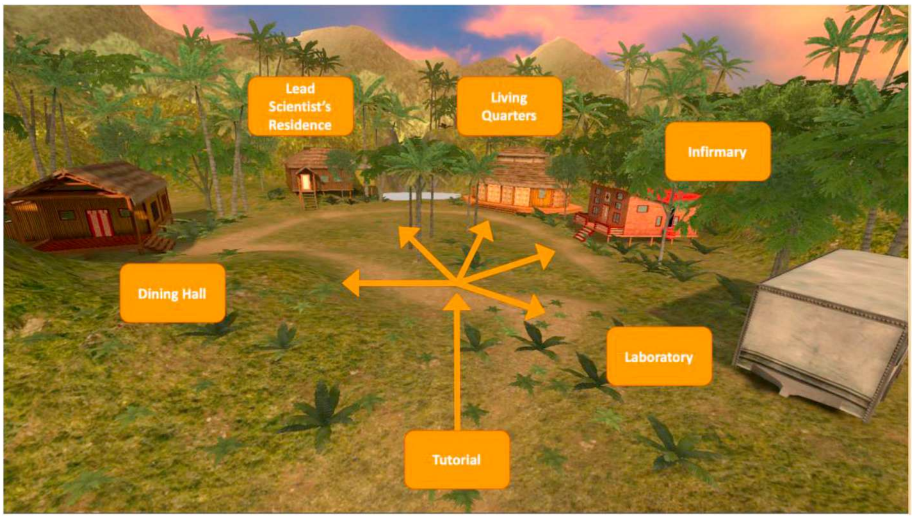

Fig. 1. Overview of the *High Agency* version of Crystal Island, where students can explore freely after completing the tutorial.

cast of characters, and microbiology content were the same as in the *Full Agency* condition. However, as depicted in Fig. 2, students were required to move between buildings in a prescribed sequence in the *Low Agency* condition. First, students completed the same gameplay tutorial as in the *Full Agency* condition. Afterward, students were required to move next to the Infirmary; they were not given the option to explore a different location or building. Unlike the *Full Agency* condition, students did not navigate the outdoor environment to move between locations. Instead, students were presented with a "fast travel" interface, which provided a menu showing available locations to which the student could teleport directly. The fast travel interface appeared on-screen whenever the student approached the exit of the current location or building. In this manner, the game imposed constraints on the order in which students explored the buildings and locations of Crystal Island by selectively making locations available or unavailable in the menu. After students finished exploring the Infirmary, they were required to move on to the Living Quarters, Lead Scientist's Residence, Dining Hall, and the Laboratory. After completing their initial tour of the five buildings, students could freely teleport between the buildings in any order they chose using the fast travel interface. In each building, students were required to comprehensively engage with every virtual character, book, object, poster, and article before exiting. This includes selecting every conversational branch in the dialog tree for each virtual character, reading every virtual book and article, and completing each embedded assessment of reading comprehension, which cumulatively imparted relevant information about microbiology concepts that were applicable toward solving the science problem scenario. If the student attempted to exit a building before completing every activity in that area, they were prompted to continue exploring the location before moving on. An example of a building containing non-player characters and books in the *Low* Agency condition is given in Fig. 3, which depicts the Infirmary in Crystal Island. In summary, student agency in this condition was restricted in two primary forms: (1) student movements between locations were pre-determined for the initial tour of the buildings, and (2) students were not allowed to progress to the next location until everything in the current location had been explored comprehensively.

Given the freedom that students in the *High Agency* condition were afforded, in-game comparisons between the two interactive conditions become more difficult due to the *structural* differences of the game versions. Structural differences are those that are specifically due to differences in game version mechanisms rather than on student cognitive, affective, metacognitive, or motivational (CAMM) processes. For example, observing that all students in the *Low Agency* condition read each book over the course of gameplay while some students in the *High Agency* condition voluntarily read each book is a structural difference since students in the *Low Agency* condition were forced to read every book in order to progress in the narrative. Since we are interested in the impact of agency on CAMM processes, the analyses were careful to isolate the structural differences due to the agency manipulation by distinguishing between three different gameplay phases (described in Section 4.3.3.) when comparing in-game problem-solving behaviors. This distinction was also made in the previous version of this work (see Sawyer et al., 2017). This allowed us to distinguish between behaviors that were required due to the nature of the condition and behaviors all students engaged in by choice. Using the same example with books, since students in the *Low Agency* condition were required to read all books and articles before progressing, it was ensured that they had all read the same number of books and articles, whereas the *High Agency* condition did not have this same structural requirement.

In the *No Agency* condition, students did not directly interact with the Crystal Island game-based learning environment. Instead, students watched a narrated video depicting the gameplay of an expert completing the Crystal Island problem scenario. All students in this condition watched the same video recording of the expert's gameplay walkthrough. The video showed the expert exploring each of the buildings in the virtual environment. The buildings were explored in the same order as the "ideal" path shown in Fig. 2. Within each building, the expert conversed with each of the virtual characters and selected every available conversational branch. The expert read each of the virtual books and articles, and in the laboratory, tested each of the potentially contaminated virtual objects. The video also depicted the expert filling out the diagnosis worksheet, identifying a diagnosis and treatment/prevention plan, and submitting the correct diagnosis to the camp nurse, thus solving the mystery. The audio narration consisted of spoken description, provided by the expert, of the actions being performed in the game. The narration did not include information about the thought process or motivations behind actions performed during gameplay. A separate study observed that students in the *High Agency* condition with similar trajectories through the problem-solving space to the expert path demonstrated marginally higher normalized learning gains (Sawyer, Rowe, Azevedo, & Lester, 2018). This agency condition represents an extreme version of restricted agency, to the point where the GBLE is not interactive since students instead watch an expert playthrough.

4. Methods

4.1. Participants and materials

The study involved 138 college students (64% female) randomly assigned to one of three agency conditions: *High Agency*, *Low Agency*, and *No Agency* (see above). This sample size was an increase from the earlier version of this study ($n = 105$). The age of students ranged from 18 to 29 ($M = 20.0$, $SD = 1.73$). Students were compensated $10/hour for participating. There were 68 students[1] in the *High Agency* condition, 38 students in the *Low Agency* condition, and 32 students in the *No Agency* condition.

The microbiology content test consisted of 21 multiple-choice questions with 12 factual (e.g., *What is the smallest type of living organism?*) and 9 application (e.g., *What is the difference between bacterial and viral reproduction?*) questions. Each question consisted of

[1] We continued data collection beyond our target of 30 per condition for the *Full Agency* condition only.

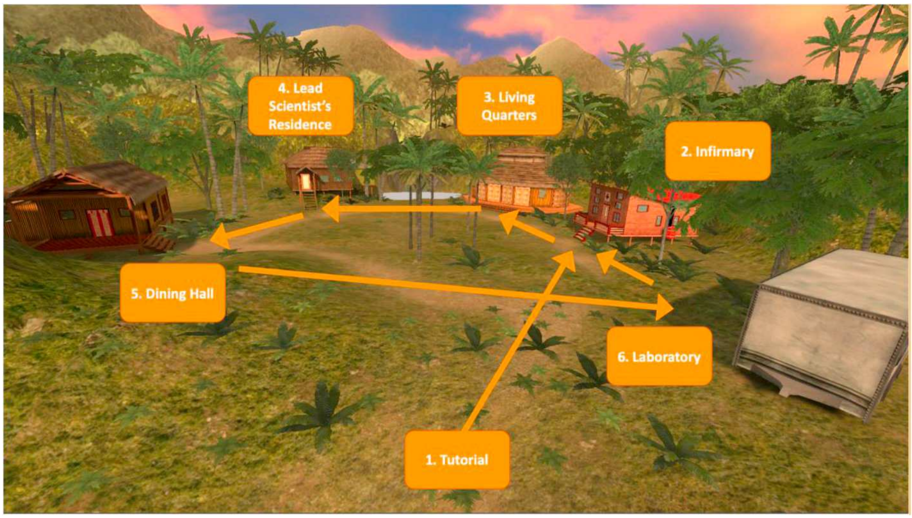

Fig. 2. The prescribed "ideal" path in Crystal Island that was provided to students in the Low Agency condition.

Fig. 3. Examples of non-player characters (NPCs) and virtual books that students in the Low Agency condition were required to interact with before progressing to the next location.

four answer choices, with one correct answer and three incorrect answers. The pre- and post-test consisted of similar but not identical questions of comparable difficulty. These were the same content tests used in the previous study. The pre-test indicated that students had limited prior knowledge of the microbiology content ($M = 12.3$ [59%], $SD = 2.8$ [13%]). A one-way ANOVA did not find evidence of a significant difference between pre-test scores among the three conditions ($F(2, 135) = 1.26$, $p = 0.29$).

In addition to the microbiology content test, student interest and engagement were assessed with several questionnaires. Student interest was assessed post-gameplay with the Intrinsic Motivation Inventory (Ryan, 1982), which has been validated across domains and consists of 29 statements in which students respond on a 7-point Likert scale describing the degree of truth for each statement (e.g. McAuley, Duncan, & Tammen, 1989). Specifically, the interest-enjoyment subscale was used as a measure of interest, which is a subset of the full assessment and consists of 7 statements that evaluate how much interest and enjoyment the student felt towards the activity. Student engagement with the virtual environment was measured after completing the game with the Presence Questionnaire (Witmer & Singer, 1998), a 33-question assessment using a 7-point Likert scale that aims to measure a student's feeling of transportation into the virtual environment. The average score after reverse-scoring the appropriate questions (8 – x, since it is a 7-point Likert) was used to get continuous values of interest and presence that range from 1 to 7. These questionnaires were a new addition from the earlier version of this study.

4.2. Experimental procedure

The same procedure used in Sawyer et al. (2017) was used for this study. Students began the session by completing a 21-question multiple-choice test assessing their conceptual and application-based understanding of microbiology, as well as several other questionnaires on emotions and motivation (not used in this study). The eye tracker and facial expression recognition technologies were then calibrated prior to gameplay. The Attention Tool software, version 6.1 (iMotions, 2016) was used for facial expression recognition. Students were then introduced to the game environment and played the game until solving the mystery or requesting to end the session, with 95% successfully solving the mystery with game durations lasting approximately 60–90 min for the interactive conditions and 91 min for the *No Agency* condition (i.e., the length of the video). Students who did not complete the mystery are included in the analysis to avoid positive impacts of selection bias on the learning and motivational outcomes. After completion of the session, students completed a counterbalanced 21-question microbiology multiple-choice test to assess learning gains and several questionnaires assessing motivational outcomes from interacting with Crystal Island.

4.3. Data coding and scoring

For this study, we used data from the content tests, self-report questionnaires, log files, and videos of facial expressions. All collected data were run through a data pipeline, which aligned all data channels. In addition, the pipeline recorded several data variables, such as in-game behaviors and self-report responses. We calculated normalized learning gain score and post-processed facial expression evidence scores. In comparison to the earlier study (Sawyer et al., 2017), the current study uses the same normalized learning gain score, but uses videos of facial expressions of emotions and self-reports of presence and interest as additional variables for this work.

4.3.1. Normalized learning gain

The scores on the microbiology pre-test and post-test were used to calculate normalized learning gain (NLG), which is used as an assessment of student learning from Crystal Island. Normalized learning gain is the difference in post- and pre-test scores standardized by the maximum possible amount of increase (for learning gains) or decrease (for learning losses) from the pre-test score of the student. This metric helps create a fair comparison among students of different measured prior knowledge, as students who scored high on the pre-test were still capable of achieving high NLG scores since their maximum possible amount of increase is comparably lower to a student who scored low on the pre-test.

$$NLG = \left\{ \begin{array}{l} \dfrac{Post - Pre}{1 - Post} \;\; if\ Post \geq Pre \\ \dfrac{Post - Pre}{Pre} \;\; if\ Post < Pre \end{array} \right\}$$

4.3.2. Facial expression evidence scores

While interacting with the game, facial expression recognition software distributed by iMotions (iMotions, 2016; previously CERT) recorded fine-grained measures of student facial expressions that correspond to the Facial Action Coding System (Ekman, 1977). This objective framework uses these action units recorded at a rate of 30 Hz to predict student emotions in real-time, giving an evidence score of the presence of several emotions. Evidence scores are the base 10 likelihood of a particular affective state being present (as would be coded by a human coder), which are produced by the framework's image-based classifiers for each affective state. This study focuses on emotions previously theorized to be relevant to the learning process: confusion, frustration, and joy (D'Mello & Graesser, 2012a). These evidence scores were preprocessed using a combination of relative and absolute thresholding of amplitude to calculate discrete events of experiencing an emotion. Specifically, after smoothing the evidence scores with an 11-step window, evidence scores were standardized by student to a unit normal distribution to account for individual differences in facial expressiveness. Thus, relative thresholding was performed by classifying an emotional event as this evidence score maintaining a standardized value above 1.65 (representing the top 5% of observations) for at least 0.5 s (to avoid micro-emotions). Absolute thresholding was performed by only classifying events as emotions if the raw evidence score was above 0.5 to avoid values that were negative (indicating the lack of the emotion) but standardized to be positive. This process yielded discrete events throughout gameplay that represent an elevated measure of emotion for at least 0.5 s. The rate of these events is used in the analysis to compare the emotional experience among agency conditions.

4.3.3. Gameplay phases

To better isolate the in-game behaviors into comparable intervals, three gameplay phases were distinguished: (1) Tutorial, (2) Information Gathering, and (3) Hypothesis Testing. As previously mentioned, these gameplay phases were examined in the previous work. However, the earlier study conceptualized the third phase as 'Diagnosis' instead of Hypothesis Testing, as in the work reported here. All versions of Crystal Island featured the same gameplay tutorial, which was presented at the start of the game for students to learn the basic game mechanics and controls (though students in the *No Agency* condition were watching the expert learn these controls). There were no differences between the *High Agency* and *Low Agency* conditions in this phase. The Information Gathering phase began immediately after completion of the tutorial, and lasted until students began conducting tests (i.e. scans) in the virtual laboratory. This phase predominately consisted of students exploring the research station and gathering information through conversations with virtual characters, as well as reading books and research articles. This phase was most different between the *High Agency* and *Low Agency* conditions due to movement restrictions and interaction requirements of the *Low Agency* condition. Once a student conducted their first test with the scanning equipment located in the virtual laboratory, the Hypothesis Testing phase began. This phase was almost identical between the *High Agency* and *Low Agency* conditions, except that in the *Low Agency* condition students utilized a fast travel interface to move between buildings, whereas in the *High Agency* condition, students walked through exterior camp locations of the 3D virtual environment. The Hypothesis Testing phase lasted until the student completed the game by successfully submitting a correct diagnosis, transmission source, and treatment/prevention plan to the camp nurse. The Information Gathering and Hypothesis Testing phases aligned closely with the model of scientific discovery as dual space search (Klahr & Dunbar, 1988), which posits that scientific reasoning occurs through two phases: the hypothesis space and experimental space. In the hypothesis space, students gather information about the given topic, and form hypotheses regarding the information they have gathered. In the experimental space, students test the hypotheses they have formed. Related to the current study, exploration in the hypothesis space was performed by the discovery and forming of hypotheses conducted in the Information Gathering phase, while exploration in the experimental space was performed by testing hypotheses through the virtual laboratory in the Hypothesis Testing phase.

Thus, within the *Low Agency* condition, the Information Gathering phase represents when student agency was restricted, and the Hypothesis Testing phase represents a period when student agency resembled the *High Agency* condition. This design allows the use of these three gameplay phases in the analysis of students' problem-solving behaviors, emotions, and agency. Specifically, two types of in-game comparisons became important after decomposing gameplay into these intervals: a between-subjects comparison and within-subjects comparison. Since both the *High Agency* and *Low* Agency conditions had similar agency restrictions in the Hypothesis Testing phase, between-subjects comparisons of in-game behaviors could be conducted. For example, we analyzed how *Low Agency* behaviors differed from *High Agency* behaviors after the agency manipulation had been applied. A within-subjects comparison was also conducted by analyzing the Information Gathering and Hypothesis Testing phases of students in the *Low Agency* condition since the Information Gathering phase had agency restrictions while the Hypothesis Testing phase did not have the same limitations on student behavior. For example, in Section 3.4, the rates of emotions of students in the *Low Agency* condition were compared between these two phases using a repeated-measures type analysis.

5. Results

For Research Questions 1 and 3, we compared students across all three conditions. Since Research Question 2 compared in-game behaviors, and students in the *No Agency* condition did not have in-game actions, we compared students from the *High* and *Low Agency* conditions for this research question. Another addition to the current paper from the previous work by Sawyer et al. (2017) is the inclusion of hypotheses. We hypothesized that:

Hypothesis 1. Students in the High Agency condition will have higher normalized learning gain scores than students in the Low or No Agency conditions.

Hypothesis 2. Students in the High Agency condition will have longer rates of scanning, reading, and conversing with non-player characters, and will have shorter rates of submitting or editing their diagnosis worksheet, than students in the Low Agency condition in the Hypothesis Testing phase of gameplay.

Hypothesis 3. Students in the High Agency condition will report higher presence, interest, and joy, and lower confusion and frustration, than students in the Low or No Agency conditions.

According to research detailing the effect of agency on learning (Bandura, 2001; Metcalfe et al., 2013; Snow et al., 2015), high levels of agency should support increases in learning outcomes and performance during game-based learning as it allows learners to maintain control over their learning experience and actively regulate their behaviors. Conversely, based on issues of discovery learning (Kirshner et al., 2006; Mayer, 2004), high levels of agency might hinder learning.

We present results to answer each of the research questions posed in the Introduction. First, results that motivate the partitioning of gameplay into the previously described gameplay phases (Tutorial, Information Gathering, and Hypothesis Testing) are reported due to differences arising from the structure of the game's design between conditions. Section 5.2 addresses RQ1 by presenting the differences in learning outcomes between conditions. Section 5.3 examines several in-game actions indicative of problem-solving behavior to answer RQ2. Section 5.4 presents results regarding the facially detected emotions from FACET and self-reported engagement and control from the Presence Questionnaire and Intrinsic Motivation Inventory, respectively, between agency conditions, answering RQ3.

5.1. Duration of gameplay

It is important to note that the amount of time spent playing the game is significantly different between conditions due to the structural differences of gameplay between conditions. A one-way ANOVA revealed significant differences in duration of gameplay between agency conditions ($F(2, 135) = 18.2$, $p < 0.001$). Since students in the *Low Agency* condition were required to interact with each object before progressing to the next location in the Information Gathering phase, students in this condition were expected to spend more time in this second phase, resulting in longer overall gameplay. The amount of duration spent in each phase by condition can visually be seen in Fig. 4 and the exact numbers are reported in Table 1. Students in the *No Agency* condition all watched the same video, resulting in the same amount of duration of gameplay for each student in the condition, which means there is no variance in duration within this condition (hence the lack of error bars for *No Agency* in Fig. 4).

5.2. Impact of agency on learning

To compare the effect of the agency condition on normalized learning gain, we conducted a one-way ANOVA. The test revealed a significant effect ($F(2, 135) = 4.79, p < 0.01$). A post-hoc analysis of the differences between conditions was conducted using a series of Welch two-sample *t*-tests, which do not assume equal population variances. In total, three pairwise tests were performed on each pair of conditions. Table 2 shows the means and standard deviations of each condition, while the normalized learning gain distribution of each condition (in terms of their density) is visualized in Fig. 5. An independent two sample *t*-test showed students in the *Low Agency* condition had marginally higher normalized learning gains than students in the *High Agency* condition ($t(81) = 1.76, p = 0.082, d = 0.35$), and significantly higher normalized learning gains than students in the *No Agency* condition ($t(64) = 3.13, p < 0.01, d = 0.76$). The remaining test showed students in the *High Agency* condition having marginally higher normalized learning gains than students in the *No Agency* condition ($t(60) = 1.81, p = 0.076, d = 0.39$). Interestingly, in the earlier study, students in the *Low Agency* condition had significantly higher NLG scores than students in the *High Agency* condition (marginal in the current study) and *No Agency* condition (as seen in this study). Similar to the previous study, there were no significant differences between the *High Agency* and *No Agency* conditions, however the significance was marginal in this current study, with a lower *p*-value than in the previous study.

5.3. Impact of agency on problem-solving behaviors

While normalized learning gain provides a high-level summary of student learning outcomes resulting from interaction with Crystal Island, another comparison of interest involves the differences in the interactions. A more fine-grained analysis of student performance in Crystal Island can be determined by examining the actions students engaged in towards solving the mystery. Differences among in-game actions between the *High Agency* and *Low Agency* conditions (the *No Agency* condition does not have any interaction with the game and thus does not have recorded in-game actions) indicate if agency impacts problem-solving behaviors. This analysis is important in assessing if problem-solving strategies differ by condition, an important consideration for encouraging effective self-regulated learning in GBLEs.

As noted above (and in the earlier study), the three agency conditions vary greatly in the amount of time students spent both overall in the game and in the Information Gathering phase, partly due to the design of the game for each condition. Thus, differences in problem-solving behavior in the Information Gathering phase are likely due to the structure of the game and do not provide insight into how the approach of students differs when their agency has been manipulated. In the Hypothesis Testing phase, students in both the *High Agency* and *Low Agency* conditions have similar restrictions on their problem-solving behaviors and have a non-significantly different amount of time spent in the interval. Therefore, a comparison among the problem-solving behaviors in this interval is most appropriate to determine how the experimental manipulation (which differs primarily in the Information Gathering phase) affects behavior in the Hypothesis Testing phase (which is similar in both conditions). To account for individual differences in duration spent in the interval, a comparison among the rate of problem-solving behaviors for each condition (count of actions divided by duration) was performed. The actions within Crystal Island related to solving the mystery that are considered problem-solving behaviors includes: scanning an item to perform a test in the virtual laboratory (Scanner), submitting the diagnosis worksheet for the final submission (Submission), reading in-game scientific books and articles (Reading), editing and taking notes on the diagnosis worksheet (Worksheet), and conversing with non-player characters (Conversation). Note that in addition to the omission of the *No Agency* condition, another five students in the *High Agency* condition were removed due to corrupted or missing game trace logs. Comparing the rate of problem-solving behaviors is another difference from the previous paper, which investigated action *counts* and *durations* separately. Results from this research question are different from the earlier version of the paper, despite the similar use of a MANOVA.

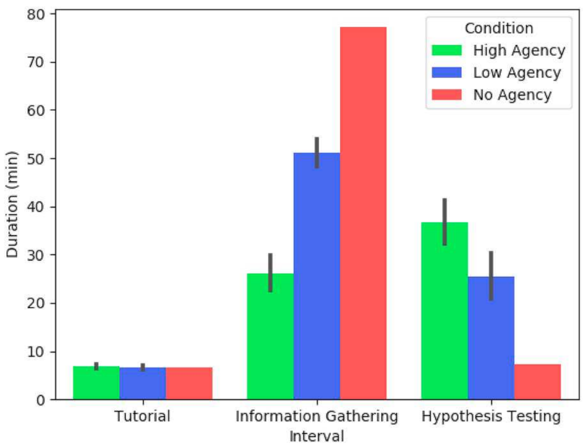

Fig. 4. Duration by gameplay phase (min) for each condition with standard error bars.

Table 1
Mean duration (min) and standard deviation by gameplay phase for each condition.

Duration Interval	High Agency	Low Agency	No Agency
Participants	$n = 68$	$n = 38$	$n = 32$
Tutorial	6.77 (1.9)	6.55 (1.3)	6.70
Information Gathering	26.0 (14.6)	51.1 (9.3)	77.1
Hypothesis Testing	36.7 (18.7)	25.5 (14.7)	7.23
All Gameplay	64.4 (27.8)	83.1 (18.2)	91.1

Table 2
Summary statistics of each condition's learning measures assessed from the content pre- and post-tests.

	Participants	Pre-Test	Post-Test	Percent Positive Learning Gain	Average Normalized Learning Gain (Std)
All Students	138	12.3	14.5	73.2%	0.256 (0.26)
High Agency	68	12.0	14.3	76.5%	0.255 (0.26)
Low Agency	38	12.1	15.3	82.6%	0.345 (0.24)
No Agency	32	12.9	14.1	56.3%	0.154 (0.26)

Note. Maximum pre- and post-test scores is 21. Percent positive learning gain = the percentage of students who earned higher post-test than pre-test scores (i.e., positive, as opposed to negative NLG score) in that condition, and overall.

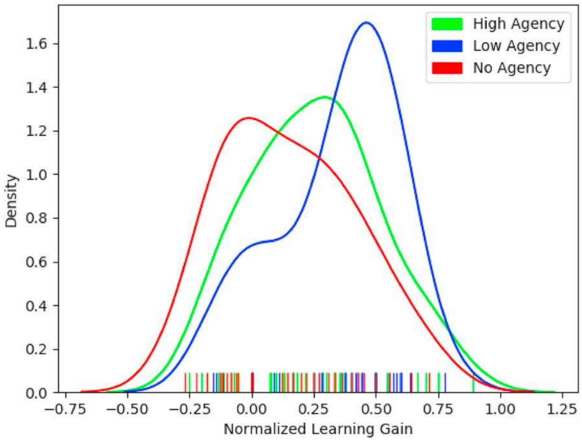

Fig. 5. Densities based on histogram of each condition's normalized learning gain observations.

A one-way MANOVA was conducted to compare the effect of agency on problem-solving behaviors of students in Crystal Island's Hypothesis Testing phase and revealed a statistically significant effect of agency on the rate of problem-solving behaviors in the Hypothesis Testing interval ($F(1, 99) = 6.85, p < 0.001$). A series of Welch's two sample *t*-tests were used to conduct a post-hoc analysis of the differences of specific problem-solving rates between *High Agency* and *Low Agency* conditions. Significance testing was performed at the $\alpha = 0.05$ level with the Holm-Bonferroni correction to account for familywise error (Holm, 1979). The results of these tests and summary statistics for the problem-solving rates of the Hypothesis Testing phase are presented in Table 3. Students in the *High Agency* condition had significantly higher rates of reading books and articles than students in the *Low Agency* condition. Students in the *Low Agency* condition had significantly higher rates of scanning objects in the virtual laboratory, submitting their worksheet, and editing their worksheet.

Table 3
Summary statistics of the rate of problem-solving behaviors in the Hypothesis Testing phase of gameplay.

	High Mean ($N = 63$)	High Std	Low Mean ($N = 38$)	Low Std	t-statistic	p-value	Cohen's d
Scanner	0.76	0.40	1.05	0.30	−4.06	<0.01*	−0.78
Submissions	0.06	0.05	0.12	0.07	−4.14	<0.01*	−0.92
Reading	0.33	0.20	0.21	0.29	2.24	0.03	0.50
Worksheet	0.54	0.30	0.76	0.56	−2.21	0.03	−0.52
Conversation	0.21	0.10	0.25	0.12	−1.81	0.07	−0.39

*Significant at the $\alpha = 0.05$ level after applying the Holm-Bonferroni correction for familywise error.

5.4. Impact of agency on emotions and engagement

An important aspect of GBLEs over traditional instructional methods is their capability to engage students with educational content so that students both learn and are engaged throughout interaction. In this section, the impact of agency on the engagement of students with the game content was measured through their learner-centered emotions and a post-survey measure of presence. The Presence Questionnaire (Witmer & Singer, 1998) is used as a measure of engagement, which specifically seeks to measure the feeling of transportation into a virtual environment experienced by students interacting with the GBLE. The Intrinsic Motivation Inventory was also given, and the Interest-Enjoyment subscale from this survey serves as a complementary measure of engagement with the GBLE across conditions. The learner-centered emotions investigated were confusion, joy, and frustration, which have been hypothesized to be related to the learning process through the model of affective dynamics (D'Mello & Graesser, 2012a). The number of times an emotion was experienced divided by total minutes of gameplay (i.e., the rate of emotion occurrences) was used to represent each student's emotional experience. Since none of these measures rely on in-game behaviors, the comparisons within this section use all three agency conditions (i.e., the *No Agency* can be included here). In comparison to the previous study, the analyses and results from this research question are a completely new addition.

A one-way MANOVA was conducted to compare the effect of agency on presence, interest-enjoyment subscale, confusion, joy, and frustration and revealed a significant effect of agency on these response variables ($F(2, 130) = 6.62$, $p < 0.001$). A post-hoc analysis using linear regression models (which give the same test statistics as one-way ANOVAs but allow for comparison of the coefficients that relate the factors to the response) for each dependent variable were calculated to determine which dependent variables were different for each condition relative to the *High Agency* condition. The results of these models are reported in Table 4 and show the average for the *High Agency*, the relative increase or decrease for both the *Low Agency* and *No Agency* conditions, and whether the differences among all groups are significant. Significance for the relative increase or decrease indicate whether that group is significantly different from the *High Agency* condition. For example, the average presence score for the *High Agency* condition was 4.99, while the *Low Agency* condition had a 0.01 higher average presence score (average of 5.0) and the *No Agency* group had a 0.94 lower average presence score (average 4.05), which was a significantly lower difference than the *High Agency* condition ("Presence" row of Table 4). This table indicates that the differences are primarily from the *No Agency* condition having lower scores on both post-survey measures of engagement and lower rates of emotions during gameplay. Post-hoc ANOVA tests comparing only the *High Agency* and *Low Agency* condition reveal no significant differences across each of the dependent variables.

Another question of interest pertains to the experience of students in the *Low Agency* condition who experienced a structured early portion of gameplay (during the Information Gathering phase) and a later portion of gameplay without restrictions on the path required to be taken through the game (during the Hypothesis Testing phase). An important investigation for future studies should examine the emotional experience of students in this condition using a within-subjects comparison between these two phases because it can directly compare students' emotions after experiencing both levels of agency. Since emotions are measured in real-time through facial expression recognition, the rates of learner-centered emotions for each student can be compared in a repeated-measures fashion over each gameplay phase. For example, one would hypothesize that students would experience a higher rate of frustration and lower rate of joy during the Information Gathering phase in which their agency was restricted compared to the Hypothesis Testing phase where they had more freedom in their action choice. A Hotelling T^2 test was conducted to compare the paired differences (between the Information Gathering phase and Hypothesis Testing phase) of rates of frustration, confusion, and joy of students in the *Low Agency* condition and revealed no significant differences between pairs of emotion rates in the two intervals ($T^2(3, 72) = 5.31$, $p = 0.17$).

6. Discussion

In this study, we expanded on an earlier study (Sawyer et al., 2017) and investigated the effect of agency on students' learning, problem-solving behaviors, emotions, and engagement (i.e., presence, interest) while interacting with a game-based learning environment that fosters SRL and scientific reasoning while being presented information about microbiology. The purpose of the study was to find further evidence of an *agency effect*, which refers to how a student's level of control or freedom making in-game decisions impacts how they demonstrate problem-solving behaviors and learning during and after gameplay, respectively. We investigated the impact of agency on students' self-regulatory behaviors, emotions, and self-reported presence and interest. According to the agency effect (Bandura, 2001; Snow et al., 2015), students with more agency during gameplay are able to exhibit more SRL behaviors, express more beneficial emotions, and express feeling more present and interested in the task, compared to students with little or no agency as these students are not able to exert control over their learning. On the other hand, based on critiques of discovery learning, high levels

Table 4
Summary of linear models with factors for each condition predicting emotions and engagement.

	High Agency mean	*Low Agency* difference	*No Agency* difference	F-statistic	R^2
Presence	4.99	0.01	−0.94**	18.75**	0.22
Interest	4.66	−0.17	−2.16**	34.94**	0.35
Confusion	1.46	0.07	−0.77*	3.03	0.045
Joy	2.06	−0.12	−1.22**	4.74*	0.068
Frustration	2.03	−0.10	−1.31**	8.15**	0.11

*$p < 0.05$, **$p < 0.01$.

of agency might hinder student learning (Kirshner et al., 2006; Mayer, 2004).

Overall, the results of this study found that agency impacted student interactions within Crystal Island in several ways, some of which are similar to the earlier version, and some of which are different. The time spent in game was significantly different among the agency conditions in both studies, primarily due to the structure of the agency conditions. The amount of learning observed was significantly different among the agency conditions, with the *Low Agency* condition achieving the largest normalized learning gain scores. This overall result was consistent across studies, but the difference was only marginal between the *Low* and *High Agency* conditions in the current study, possibly due to an increase in sample size. Finally, as investigated only in this study, results revealed differences in emotion and engagement outcomes among agency conditions, but surprisingly no distinguishable differences between the *High Agency* and *Low Agency* conditions were observed.

6.1. Agency and duration and relationship with learning

The time of gameplay for each student depended on how long the student took to solve the mystery, meaning there were differences among the durations of gameplay for students in the study (for both studies). The differences in duration between conditions are to be expected given the structural nature of the agency conditions, and they drive the context for the analysis of the other variables of interest. However, the differences in duration have important implications for implementing agency manipulations in run-time environments. The current study was conducted in a laboratory setting in which students were provided as much time as they needed to solve the mystery. If the GBLE was to be deployed under different settings, then the increased amount of time in the *Low Agency* and *No Agency* conditions would be an important consideration under the time constraints present in alternative environments, such as classrooms. Since students spend significantly more time in the game in the *Low Agency* condition, a run-time environment, which adaptively selects between conditions should be aware of the total amount of time allocated to playing the game in addition to other cognitive, affective, metacognitive, and motivational effects of agency manipulations. For example, the game could scaffold how to engage in metacognitive monitoring strategies to keep track of time and which items students should be interacting with that are relevant to solving the mystery. The game could also foster engaging in cognitive learning strategies, such as reading books in an efficient way (e.g., Taub et al., 2017) or testing food items efficiently, as opposed to testing everything (e.g., Taub & Azevedo, 2018; Taub et al., 2018). This practical consideration becomes especially important in contexts such as classrooms, which have only the length of the class period to allow students to interact with the GBLE.

Given the significant differences in duration between agency conditions, the learning analysis focused on normalized learning gain rather than a standardized measure of learning by time played because the experiment controlled for agency and had no duration restriction, meaning that agency impacted duration. Thus, mathematically, the duration (or time on task) is not independent of agency manipulation, and any learning rates calculated using these durations will be dependent on agency manipulation, essentially introducing a confounding variable to the agency-learning analysis. In these tests of learning, we wish to isolate the effect of agency on learning, i.e., to test whether learning is independent of agency manipulation. This effect would essentially be guaranteed once the impact of agency on duration is established if the rates of learning are compared. In other words, we wish to avoid a false positive from a significant difference in learning rate due to the underlying variable of duration, which is caused by the gameplay structure of the agency conditions. The analysis conducted in Sections 5.1 and 5.2 are analogous to post-hoc tests of a MANOVA for the effect of agency condition on normalized learning gain and duration, which yielded significant results ($F(2,130) = 10.2, p < 0.001$). The analysis is not explicitly reported in the results section since the *No Agency* condition lacking variance in duration practically guarantees significant differences at this granularity of test.

To alleviate concerns that the duration in game (time on task) causes the difference in learning gain, separate tests were conducted to determine if learning is independent of duration within each group. The purpose of these analyses is to determine if the time on task and learning gains are independent within each agency condition. If they are independent, then the choice of using normalized learning gain as the measure of learning is justified over using a time-standardized measure of learning (such as learning rate) in the context of this experiment. Each group was analyzed independently to prevent agency from having a confounding effect on learning. Neither the *Low Agency* nor *High Agency* groups had significant correlations between normalized learning gain and duration ($r = -0.14$, $p = 0.41$ for *Low Agency*; $r = 0.04$, $p = 0.75$ for *High Agency*). Since correlation is only a measure of the linear relationship between two variables, linear models predicting normalized learning gain from polynomial-transformed duration were also conducted within groups to test for higher order relationships between the two variables. Polynomial feature augmentation up to a degree-10 polynomial was performed on the duration variable to increase the model capacity to predict normalized learning gain from duration. For example, a degree three polynomial in a linear regression model would estimate coefficients β for $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$ with $x =$ duration of gameplay, and $y =$ normalized learning gain. The significance of this model would be tested using the traditional nested *F*-test. For each group, none of the 10 linear models created (one for each polynomial transformation up to a degree-10 polynomial, where degree $= 1$ yields the same results as a correlation test) were significant at a lenient $\alpha = 0.1$ level. While not completely exhaustive, this analysis indicates that according to the data observed in this study, $p(L \mid D, C) = p(L \mid C)$, or that normalized learning gain given agency condition is conditionally independent of duration. Given this property, the test of learning given agency condition performed in section 5.2 is most appropriate to determine if learning is independent of agency condition, as opposed to some time-standardized measure of learning. Since there is no variance in the *No Agency* condition student durations, any variation in learning is independent of duration for this condition.

6.2. Agency and learning and problem-solving

The primary high-level takeaway of the experiment is that agency had an effect on the normalized learning gain of students, as indicated by the results from Section 5.2. Specifically, students in the *Low Agency* condition had significantly higher normalized learning gains than students in the *No Agency* condition, including a large effect size of Cohen's $d = 0.76$. While not significant at the $\alpha = 0.05$ level, the comparisons with the *High Agency* condition had medium effect sizes such that an ordering of learning from minimum learning to maximum learning would be: *No Agency* < *High Agency* < *Low Agency*. Despite not revealing the same significance level, this minimum to maximum learning trend was found in both studies. Therefore, in the current study, interpretations can be made regarding the benefit of providing *some* level of agency to students during game-based learning, as opposed to providing no agency at all.

These results, therefore, do not support Hypothesis 1 because we predicted a significantly higher NLG score for students in the *High Agency* condition. However, NLG scores were not significantly different between *High* and *Low Agency* conditions; and in addition, mean NLG scores were higher for students in *the Low Agency* condition (although not significantly higher). This supports previous work by Nguyen et al. (2018) who also did not find differences in learning between high vs. low agency conditions. Additionally, students in the *No Agency* condition had significantly lower NLG scores than students in the *Low Agency* condition, which does not confirm Hypothesis 1 either.

These findings do support previous work on discovery learning (Kirshner et al., 2006; Mayer, 2014) that suggest there are benefits to sacrificing a degree of student agency by providing guidance to ensure that students engage with all instructional content that is available. In Crystal Island, the *Low Agency* condition is required to interact with all books and articles as well as speak with each non-player character guaranteeing that students in this condition experience all scientific content available. Since the content post-test is related to the content within these scientific articles, the manner in which the agency was restricted appears to be effective in encouraging learning through interacting with all scientific content. This is contrasted by the *High Agency* condition, which does not require students to necessarily experience all science content, and which yielded marginally lower learning gains, supporting the discovery learning hypothesis in this GBLE.

From an SRL perspective, these results suggest that students in the *Low Agency* condition spent more time engaging in learning and information processing. However, they were not required to metacognitively monitor which items to interact with, suggesting that the higher rate in the Hypothesis Testing phase is indicative of engaging in more cognitive learning strategies, resulting in higher NLG scores than students who were not able to engage in these learning strategies because they were only watching a video playthrough of the game.

However, while the *Low Agency* condition appeared to learn more than the *High Agency* condition, the analysis of problem-solving behaviors by gameplay phase indicates that there may be a tradeoff in addition to the increased duration. In the phase in which students had similar restrictions (the Hypothesis Testing phase) the *Low Agency* students exhibited a higher propensity for "guess-and-check" problem-solving behaviors. Specifically, the *Low Agency* condition had significantly higher rates of scanning items and submitting worksheets as their final submissions compared to the *High Agency* condition. These results partially support Hypothesis 2. We predicted students in the *High Agency* condition would exhibit higher rates of scanning, which we did not find. However, we also predicted lower rates of editing and submitting the diagnosis worksheet and higher rates of reading, which we did find. Since the game completes upon a successful submission, this indicates these students had more incorrect solution proposals than students in the *High Agency* condition. Meanwhile, the students in the *High Agency* condition had a higher rate of reading books and articles in this same phase. This could be because students in the *Low Agency* condition had read the content already and were hesitant to go re-read books and articles they had already seen. However, when considering that *Low Agency* students performed more "guess-and-check" behaviors, the lack of re-reading indicates ineffective self-regulated learning since students were not adapting their problem-solving strategies but instead were performing more incorrect scans and solution submissions. Students in the *Low Agency* condition were not required to engage in metacognitive monitoring strategies during reading as they were required to read all content, thereby they did not need to monitor what they were reading or select cognitive learning strategies. This lack of engaging in SRL during this phase might have therefore impacted their ability to self-regulate in the subsequent phase because they were not engaging in this skill. Thus, in this condition, there was a tradeoff between agency and SRL: students had higher learning gains because they did not have to engage in SRL and select what to read, but this also impeded their ability to engage in SRL at other instances. Thus, these results align with Snow et al. (2015) who found students with high levels of agency engaged in better self-regulatory processes (i.e., self-explanation) because it seems as though students in the *High Agency* condition were able to monitor their behaviors, thus having lower rates of scanning (i.e., guessing and checking) and submitting the worksheet (i.e., guessing the solution), indicative of less maladaptive self-regulatory behaviors. We also predicted students in the *High Agency* condition would have higher rates of reading activities because they would have engaged in them less during the Information Gathering phase, but granting them high levels of agency also granted them the ability to engage in more effective self-regulatory processes.

6.3. Agency and emotions and interest

The theoretically proposed tradeoff of agency includes what was proposed in Hypothesis 3: Increasing restrictions on agency leads to more negative emotions and less engagement in scientific content. This trend was partially supported by the results from Section 5.4, which found that the *No Agency* condition reported significantly lower presence and interest with the GBLE. However, outside of this extreme restriction on agency (a non-interactive version of the game in the *No Agency* condition), no differences were observed among rates of emotions, presence, or interest between the *High Agency* and *Low Agency* conditions. Further, students in the *Low Agency* condition did not experience different rates of emotions in the more restrictive phase of gameplay compared to the later phase of gameplay with less restrictions. Therefore, these results do not support Hypothesis 3, as the *Low Agency* condition had no significant differences from the *High Agency* condition among these measured outcomes, and we predicted students in the *High Agency* condition would have higher reported presence and interest, and higher evidence of joy, and lower evidence of confusion and frustration. This aligns with Nguyen et al. (2018) who did not find differences in engagement between high and low levels of agency. However, we did find a significantly higher report of engagement for students in the *High Agency* condition compared to students in the *No Agency* condition, which supports Hypothesis 3, but does not align with Veinott et al. (2013) who did not find any differences in engagement between students with agency versus without agency.

Since the overall study uses a between-subjects design, the lack of differences could be due to students only receiving one treatment. For example, if students who played the *High Agency* version then played the *Low Agency* version of the game, they may have been more frustrated and less interested since the reduction in agency was now apparent. In other words, the agency manipulation was hidden to the students in the *High Agency* and *Low Agency* groups, which could have prevented negative emotions and engagement in the *Low Agency* groups. Supporting this hypothesis was the *No Agency* condition, in which students watched an expert play through the game. In this condition, the agency manipulation was more visible since students were observing another individual play through the *High Agency* version of the game. In other words, students in the *Low Agency* condition may not have known what they were missing from an agency standpoint while students in the *No Agency* condition were observing what a higher agency version of the game was, and thus had lower engagement outcomes. Given the lower rates of emotion observed in the *No Agency* condition, another explanation for the lower interest and engagement is that this version of the game was less stimulating for students. However, given the current study, it is not possible to determine the validity of this hypothesis about the lack of different engagement outcomes, as it would require a dedicated within-subjects study to determine the effect of mixing the agency conditions for one student. This is an important consideration for implementing agency manipulations at run-time, as careful consideration will have to be given to restricting student agency at run-time after they have experienced a higher agency version of the game.

Theoretically, these results suggest that the lack of agency did not allow for students to engage in SRL processes, thereby not allowing them to experience any presence in the game or interest in the task. In addition, students who did not have any agency were not given the opportunity to experience an impasse or try to resolve confusion or frustration, resulting in lower evidence scores of these emotions. They likely did not enjoy this lack of opportunity, thus not expressing joy either. The lack of significant differences between having high or low agency might indicate that as long as students have some form of agency while interacting with the game, they will not experience differences in presence, interest, or learner-centered emotions, even though the level of agency did impact the rate of engaging in maladaptive SRL behaviors. These results emphasize the importance of granting students some level of agency during learning with GBLEs as it will impact their overall learning, engagement, and emotions.

6.4. Implications of the agency effect

Results from this study have important implications for student learning and the degree of agency afforded to students within a game. First, our results reveal that although research has demonstrated students with greater agency will outperform students with less agency in a learning task, we must also consider how different *levels* of agency play an important role on learning, performance, and affect. For example, providing students with agency (compared to no agency at all) is beneficial (Snow et al., 2015), and this was confirmed by the current study. However, our results also indicated that students who were given low levels of agency obtained the highest normalized learning gain score, not the students who were given the highest amount of agency. In contrast, students with more agency had lower rates of scanning objects and proposing a solution, demonstrating more efficient gameplay. Furthermore, there were no observed significant differences in expressed emotions or self-reported interest or presence between the high and low agency conditions, which suggests that high and low levels of agency impact students' affect similarly. Students who were given no agency had the lowest amounts of all affective responses. Thus, our results demonstrate that the level of agency can be important for impacting learning gain and performance, but providing any type of agency (compared to no agency) is beneficial for engagement and interest. Therefore, it appears that degree of agency can have a different impact on different types of actions or behaviors.

Additionally, our results inform us that providing agency is beneficial, but we do not yet know *when* this is the case. For example, is it beneficial to provide low levels of agency, followed by high levels of agency? Or should agency be induced, supported, or faded based on accuracy of self-regulatory and scientific reasoning behaviors during gameplay? In the current study, even when students were provided with low agency, they tended to engage in higher rates of activities (scanning, worksheet submissions, worksheet edits) in the hypothesis testing phase, which are seen as less efficient behaviors (Taub et al., 2018). Thus, it is still unclear when providing different levels of agency might be the most beneficial for students.

Other factors that might influence the impact of different types of agency might be internal to the student. This can include prior domain knowledge, cognitive ability, goal orientation, or other motivational factors. In this study, interest was self-reported significantly less in the *No Agency* condition. Given that in this condition students do not even play the game, it is likely that motivational factors will remain low. As one of the main goals of GBLEs is to maintain high levels of engagement and motivation during learning (Mayer, 2014), it appears that providing no agency at all will threaten this purpose.

With regards to cognitive processes and mental workload, research has shown that students' level of prior domain knowledge has been found to impact their use of SRL strategies (Taub & Azevedo, 2019; Taub, Azevedo, Bouchet, & Khosravifar, 2014), which can be attributed to their level of cognitive load. Specifically, if students have high domain knowledge, they should have the cognitive capacity to engage in cognitive and metacognitive processes. However, if students have low prior knowledge, they must allocate enough resources to focus on learning the content material, giving them fewer resources to engage in these higher order processes. Thus, it may be the case that students with different levels of domain knowledge benefit differently from different levels of agency. Perhaps students with lower prior knowledge need more guidance in choosing content to read. Therefore, lower levels of agency will provide them with the guidance they need. In contrast, students with high prior knowledge do not need any restrictions on choosing content to read, and would thus benefit more from higher levels of agency. On the other hand, in the current study, we found that students with low agency were less efficient during the Hypothesis Testing phase, which suggests these students were not self-regulating as effectively as students with high agency, despite having read all the content in the Information Gathering phase and presumably acquiring the necessary domain knowledge. Future research should include prior domain knowledge as an independent measure to determine if this effect occurs.

Level of agency can also have an impact on students with different ability levels. For example, if students have a learning disability, this impacts their cognitive functioning (e.g., memory), which limits their capabilities in self-regulation (Mason & Reid, 2018). It is unclear, however, how such students engage in self-regulatory processes during learning with advanced learning technologies. There have been many interventions that implement the use of technology (such as cell phones and iPads) to improve self-regulatory and on-task behaviors such as self-monitoring (Mason & Reid, 2018). However, such interventions have not included game-based learning. It is therefore unclear how providing different levels of agency would impact students with special needs, since a game like Crystal Island requires engaging in multiple activities. For example, students must choose which location to navigate to and which objects to interact with, in addition to completing the embedded assessments and filling out the diagnosis worksheet. This can pose a possible threat to cognitive overload, especially if a student has cognitive deficits. Therefore, perhaps these students could benefit from lower levels of agency because they would not be required to use an overload of cognitive resources that would be required in the *High Agency* condition. Future studies are needed to investigate this relationship.

6.5. Limitations

Although results indicated significant differences between agency conditions, which will inform future studies on the design of GBLEs, there are several limitations that should be acknowledged. First, there were only three conditions of student agency that were developed as versions of the GBLE, which makes claims regarding the full spectrum of agency difficult to justify. Therefore, we can only make claims regarding the relative increase or decrease of agency from the examined conditions instead of more generalizable claims about agency in games. Second, learning was an important response variable in this study, but it was measured using a pre- and post-test, which is an inherently noisy measure of learning immediately succeeding gameplay. The nature of this assessment means this measure does not assess deep learning related to long-term retention of concepts (Graesser, 2017), a potential benefit of immersive game-based learning over traditional instructional methods (Gee, 2003). Third, the measures of engagement came from a post-gameplay survey, which requires students to reflect on their entire gameplay at the conclusion of their session. Thus, measures of engagement during play, such as students entering a state of flow (Hamari et al., 2016), are difficult to determine from the measures used in this study. Fourth, the in-game measures of emotions rely on the accuracy of the evidence scores provided by iMotions. While these scores have been empirically validated (Dente, Küster, Skora, & Krumhuber, 2017) and have been found to be of use in predicting retrospective judgments of confidence using the same preprocessing as in this study (Sawyer, Mudrick, Azevedo, & Lester, 2018), they still provide a source of uncertainty that should be acknowledged.

7. Conclusion and future work

In this paper, we examined how different levels of agency impacted students' learning, emotions, and self-reported engagement and presence during learning with a game-based learning environment, Crystal Island. Previous research describes agency as providing students with autonomy during learning, where providing agency leads to better learning outcomes (Bandura, 2001; Snow et al., 2015).

Few studies have investigated the impact of *different levels* of agency on students' learning outcomes, problem-solving behaviors, emotions, and engagement, which was the goal of this study, where a previous study investigated the impact of levels of agency on learning and problem solving only (Sawyer et al., 2017). We developed three experimental conditions differing in the level of agency provided to students: *High*, *Low*, and *No Agency*. Results revealed significantly higher learning gains for students in the *Low* vs. *No Agency* condition and marginally significantly higher learning gains for students in the *Low* vs. *High Agency* conditions. Additionally, students in the *No Agency* condition expressed lower levels of joy, confusion, and frustration and self-reported lower levels of engagement and interest. There were no differences between the *High* and *Low Agency* conditions in these affective measures.

These results suggest that learning is the most effective when students are provided with *some* level of agency compared to having no agency at all. It appears that providing agency, but still restricting students in some way seems to be the most beneficial. In terms of game-based learning, findings from this study reveal that some level of scaffolding can be helpful for students, and does not compromise their levels of interest and engagement, which is a main goal of GBLEs.

As discussed in Section 6.4., a potential explanation for the lack of differences in emotional and engagement measures between the *High Agency* and *Low Agency* conditions could stem from the between-subjects design of the agency manipulation (i.e., students are not aware that they could be given more autonomy during learning). Therefore, a promising direction for future work is to conduct a within-subjects design of agency manipulation to determine if the agency effect holds true when students are provided with different levels of agency within a learning session. In this design, students would be randomly assigned to either the *High Agency* or *Low Agency* condition during the first half of gameplay and the opposite condition during the second half of gameplay. This would cause the agency manipulation to be more apparent for students, which would likely draw more polarizing reactions as students acknowledge the differences between agency versions. Thus, the hypothesis would be that students rate the *Low Agency* version of the game as less engaging and generally experience more frustration during this condition, as opposed to being provided more autonomy during learning. Additionally, this would allow us to measure and examine how SRL knowledge and skills are used across agency conditions. Another interesting component for future research would be to implement a reflective phase in-between the switch from one agency condition to another.

Importantly, future work should consider alternate data channels and modalities for assessing students during gameplay. For example, eye tracking could help inform whether reading strategies or off-task behavior differs between agency conditions. Galvanic skin response data could provide a complementary measure of engagement or frustration. Incorporating this multimodal approach would benefit the analysis by providing a more comprehensive view of a student's experience during gameplay in the different agency conditions (see Azevedo et al., 2018, 2019).

A related area of future research would be the development and assessment of a GBLE that adapts agency at run-time based on the gameplay and knowledge of the student. Based on the knowledge from this study, a basic rule for adaptation could be to assign the *Low Agency* condition to students who appear to be off-task or who have not interacted with relevant scientific content. Ideally, a data-driven methodology such as training an interactive narrative planner with reinforcement learning (e.g. Rowe & Lester, 2015; Sawyer, Rowe, & Lester, 2017) would determine the optimal policy for agency assignment during run-time.

7.1. Conclusion

A key purpose of game-based learning is to provide rich, interactive learning experiences that are simultaneously effective and engaging for students. Game-based learning environments are often designed to provide students with significant agency to explore and solve problems in a manner of their choosing. The freedom to pursue tasks based on students' own personal preferences can lead to increased student interest and engagement while preventing a negative emotional experience. However, the greater freedom that is provided, the less structure that is available, sometimes leading to reduced learning outcomes.

To test the effect of student agency in GBLEs, we conducted a complementary study to one reported in an earlier paper, where both studies assigned students to three conditions for interacting with the Crystal Island game environment: a *High Agency* condition, a *Low Agency* condition, and a *No Agency* condition. The observed results partially support previous agency hypotheses since the *High Agency* condition achieved marginally lower normalized learning gains compared to the *Low Agency* condition. However, the *No Agency* condition yielded the lowest learning gain, demonstrating that extreme agency restrictions did not provide learning benefits. While the learning results generally followed previous hypotheses on agency, the tradeoff that reduced agency is associated with reduced amounts of interest and engagement was not found between interactive agency conditions. More specifically, in this study no differences between the interactive versions of the game, *Low Agency* and *High Agency*, were observed comparing rates of confusion, frustration, and joy measured through facial expression recognition in real-time, and measures of interest and presence from post-surveys. The *No Agency* condition reported significantly lower interest and presence with less rates of confusion, frustration, and joy, indicating an overall less stimulating experience with the GBLE. Thus, there were no observed detriments to the *Low Agency* condition from measured interest, presence, and rates of negative or positive emotions during learning.

The results suggest that extreme restrictions on agency (i.e., the *No Agency* condition) can be detrimental to learning and engagement outcomes through providing a less stimulating experience. The results also suggest that moderate restrictions on agency (the *Low Agency* condition) can result in increased learning, yet ineffective SRL, but not cause decreased interest or engagement relative to an unrestricted agency version of the game (the *High Agency* condition). These results have important implications for the design of GBLEs that target increased learning outcomes and increased use of effective self-regulatory processes, as agency is a central design consideration for these environments.

Acknowledgments

We would like to thank our collaborators in the SMART Lab at the University of Central Florida and the Center for Educational Informatics at North Carolina State University. The authors would like to specifically thank Robert Taylor for his assistance with the system and data pipeline development and Elizabeth Cloude for assistance with data collection. This study was supported by funding from the Social Sciences and Humanities Research Council of Canada (SSHRC 895–2011–1006). Any conclusions expressed in this material do not necessarily reflect the views of SSHRC.

References

Azevedo, R., Mudrick, N. V., Taub, M., & Bradbury, A. E. (2019). Self-regulation in computer-assisted learning systems. In J. Dunlosky, & K. Rawson (Eds.), *The Cambridge handbook of cognition and education* (pp. 587–618). Cambridge, MA: Cambridge Press.

Azevedo, R., Taub, M., & Mudrick, N. V. (2018). Using multi-channel trace data to infer and foster self-regulated learning between humans and advanced learning technologies. In D. H. Schunk, & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 254–270). New York, NY: Routledge.

Baker, R., Clarke-Midura, J., & Ocumpaugh, J. (2016). Towards general models of effective science inquiry in virtual performance assessments. *Journal of Computer Assisted Learning, 32*(3), 267–280.

Baker, R., Moore, G., Wagner, A., Kalka, J., Salvi, A., Karabinos, M., et al. (2011). The dynamics between student affect and behavior occurring outside of educational software. In *Proceedings of the 4th international conference on affective computing and intelligent interaction* (pp. 14–24).

Baker, R. S. J. D., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies, 68*(4), 223–241.

Bandura, A. (2001). Social cognitive theory : An agentic perspective. *Annual Review of Psychology, 52*, 1–26.

Barab, S., Dodge, T., Jackson, C., & Arici, A. (2003). *Technical report on quest Atlantis*.

Barab, S., Thomas, M., Dodge, T., Carteaux, R., & Tuzun, H. (2005). Making learning fun: Quest Atlantis, a game without guns. *Educational Technology Research & Development, 53*(1), 86–107.

Calvert, S. L., Strong, B. L., & Gallagher, L. (2005). Control as an engagement feature for young children ' s attention to and learning of computer content. *American Behavioral Scientist, 48*(5), 578–589.

Calvo, R., D'Mello, S., Gratch, J., & Kappas, A. (2015). *The Oxford handbook of affective computing*. New York, New York, USA: Oxford University Press.

Clark, D., Tanner-Smith, E., & Killingsworth, S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research, 86* (1), 79–122.

Dente, P., Küster, D., Skora, L., & Krumhuber, E. G. (2017). Measures and metrics for automatic emotion classification via FACET. In *Proceedings of the conference on the study of artificial intelligence and simulation of behaviour (AISB)* (pp. 160–163).

Domagk, S., Schwartz, R., & Plass, J. (2010). Interactivity in multimedia learning: An integrated model. *Computers in Human Behavior, 26*, 1024–1033.

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22*(2), 145–157.

D'Mello, S., & Graesser, A. (2012). AutoTutor and Affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back? *ACM Transactions on Interactive Intelligent Systems, 15*(212), 434–442.

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction, 29*, 153–170.

Easterday, M., Aleven, V., Scheines, R., & Carver, S. (2016). Using tutors to improve educational games: A cognitive game for policy argument. *The Journal of the Learning Sciences*, 1–51.

Ekman, P. (1977). *Facial action coding system*. Mountain View, CA: Consulting Psychologists Press.

Ekman, P. (1984). Expressions and the nature of emotion. In K. Scherer, & P. Ekman (Eds.), *Approaches to emotion* (pp. 319–344). Hillsdale, NJ: Erlbaum.

Fredricks, J., Blumenfeld, P., & Paris, A. (2004). School engagement: Potential of the concept, state of the art evidence. *Review of Educational Research, 74*, 59–109.

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment, 1*(1), 20.

Graesser, A. C. (2017). Reflections on serious games. In H. van Oostendorp, & P. Wouters (Eds.), *Instructional techniques to facilitate learning and motivation of serious games* (pp. 199–212). AG, Switzerland: Springer.

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior, 54*, 170–179.

Harp, S., & Mayer, R. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology, 90*(3), 414–434.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*, 65–70.

iMotions. (2016). *Attention tool* (version 6.1) [Computer software]. Boston, MA: iMotions Inc.

Kim, H., & Ke, F. (2017). Effects of game-based learning in an OpenSim-supported virtual environment on mathematical performance. *Interactive Learning Environments, 25*(4), 543–557.

Kirshner, P., Sweller, J., & Clark, R. (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*(2), 75–86.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*(1), 1–48.

Long, Y., & Aleven, V. (2014). Gamification of joint student/system control over problem selection in a linear equation tutor. In *Proceedings of the 12th international conference on intelligent tutoring systems* (pp. 378–387).

Malone, T., & Lepper, M. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. *Aptitude, Learning, and Instruction, 3*, 223–253.

Mason, L. H., & Reid, R. (2018). Self-regulation: Implications for individuals with special needs. In D. H. Schunk, & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (*2nd* ed., pp. 473–484). New York, NY: Routledge.

Mayer, R. (2004). Should there be a three-strikes rule against pure discovery learning? The case for guided methods of instruction. *American Psychologist, 59*(1), 14–19.

Mayer, R. (2014). *Computer games for learning: An evidence-based approach*. MIT Press.

McAuley, E., Duncan, T., & Tammen, V. (1989). Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise & Sport, 60*(1), 48–58.

Metcalfe, J., Eich, T. S., & Miele, D. B. (2013). Metacognition of agency: Proximal action and distal outcome. *Experimental Brain Research, 229*(3), 485–496.

Mott, B. W., & Lester, J. C. (2006). Narrative-centered tutorial planning for inquiry-based learning environments. In *Proceedings of the 8th international conference on intelligent tutoring systems* (pp. 1–10). ITS-2006.

Nguyen, H., Harpstead, E., Wang, Y., & McLaren, B. M. (2018). Student agency and game-based learning: A study comparing low and high agency. In *Proceedings of the 19th international conference on artificial intelligence in education* (pp. 338–351). Springer International Publishing.

Nietfeld, J. L. (2018). The role of self-regulated learning in digital games. In D. H. Schunk, & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (*2nd* ed., pp. 271–284). New York, NY: Routledge.

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist, 50*(4), 258–283.

Rowe, J. P., & Lester, J. C. (2015). Improving student problem solving in narrative-centered learning environments : A modular reinforcement learning framework. In *Proceedings of the 17th international conference on artificial intelligence in education* (pp. 419–428).

Rowe, J., McQuiggan, S., Robison, J., & Lester, J. (2009). Off-task behavior in narrative-centered learning environments. In *Proceedings of the 14th international conference on artificial intelligence in education* (pp. 99–106). Brighton: IOS Press.

Rowe, J. P., Shores, L. R., Mott, B. W., & Lester, J. C. (2011). Integrating learning, problem solving, and engagement in narrative-centered learning environments. *International Journal of Artificial Intelligence in Education, 21*(1–2), 115–133.

Ryan, R. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology, 43*(3), 450–461.

Ryan, R., & Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, 68–78.

Sabourin, J. L., & Lester, J. C. (2014). Affect and engagement in game-based learning environments. *5*(1), 45–56.

Sao Pedro, M., Baker, R., Gobert, J., Montalvo, O., & Nakama, A. (2013). Leveraging machine-learned detectors of systematic inquiry behavior to estimate and predict transfer of inquiry skill. *User Modeling and User-Adapted Interaction, 23*, 1–39.

Sawyer, R., Mudrick, N. V., Azevedo, R., & Lester, J. (2018). Impact of learner-centered affective dynamics on metacognitive judgements and performance in advanced learning technologies. In *Proceedings of the 19th international conference on artificial intelligence in education* (pp. 312–316). Springer International Publishing.

Sawyer, R., Rowe, J., & Lester, J. (2017). Balancing learning and engagement in game-based learning environments with multi-objective reinforcement learning. In *Artificial intelligence in education* (pp. 323–334).

Sawyer, R., Smith, A., Rowe, J., Azevedo, R., & Lester, J. (2017). *Is more agency better? The impact of student agency on game-based learning* (pp. 335–346). Amsterdam, The Netherlands: Springer.

Sawyer, R., Rowe, J., Azevedo, R., & Lester, J. (2018). *Filtered time series analyses of student problem-solving behaviors in game-based learning* (pp. 229–238). Buffalo, NY: Educational Data Mining Society.

Scalise, K., & Clarke-Midura, J. (2014). mIRT-bayes as hybrid measurement model for technology-enhanced assessments. In *Annual meeting of the national council of measurement in education. Philadelphia, PA*.

Shute, V. J., Rahimi, S., & Lu, X. (2019). Supporting learning in educational games: Promises and challenges. In P. Diaz, A. Ioannou, K. K. Bhagat, & J. M. Spector (Eds.), *Learning in a digital world - perspective on interactive technologies for formal and informal education* (pp. 59–81). New York, NY: Springer.

Snow, E. L., Allen, L. K., Jacovina, M. E., & McNamara, D. S. (2015). Does agency matter?: Exploring the impact of controlled behaviors within a game-based environment. *Computers and Education, 82*, 378–392.

Taub, M., & Azevedo, R. (2018). Using sequence mining to assess self-regulated learning and scientific inquiry based on levels of efficiency and emotional expressivity during game-based learning. *Journal of Educational Data Mining, 10*, 1–26.
Taub, M., & Azevedo, R. (2019). How does prior knowledge influence eye fixations and sequences of cognitive and metacognitive SRL processes during learning with an intelligent tutoring system? *International Journal of Artificial Intelligence in Education, 29*(1), 1–28. https://doi.org/10.1007/s40593-018-0165-4.
Taub, M., Azevedo, R., Bouchet, F., & Khosravifar, B. (2014). Can the use of cognitive and metacognitive self-regulated learning strategies be predicted by learners' levels of prior knowledge in hypermedia-learning environments*? Computers in Human Behavior, 39*, 356–367.
Taub, M., Azevedo, R., Bradbury, A. E., Millar, G. C., & Lester, J. (2018). Using sequence mining to reveal the efficiency in scientific reasoning during STEM learning with a game-based learning environment. *Learning and Instruction, 54*, 93–103.
Taub, M., Azevedo, R., Bradbury, A. E., & Mudrick, N. V. (2019). The importance of self-regulation and reflection during game-based learning. In J. L. Plass, B. D. Homer, & R. E. Mayer (Eds.), *Handbook of game-based learning*. Boston, MA: MIT Press (in press).
Taub, M., Mudrick, N. V., Azevedo, R., Millar, G. C., Rowe, J., & Lester, J. (2017). Using multi-channel data with multi-level modeling to assess in-game performance during gameplay with CRYSTAL ISLAND. *Computers in Human Behavior, 76*, 641–655.
Veinott, E. S., Leonard, J., Lerner, E., Perelman, B., Hale, C., Catrambone, R., et al. (2013). The effect of camera perspective and session duration on training decision making in a serious video game. In *IEEE international games innovation conference* (pp. 256–262).
Ventura, M., Shute, V., & Kim, Y. (2013). Assessment and learning of qualitative physics in Newton's playground. *Journal of Educational Research, 106*, 423–430.
Wardrip-Fruin, N., Mateas, M., Dow, S., & Sali, S. (2009). Agency reconsidered. In *Proceedings of digital games research association 2009*.
Winne, P. (2018). Cognition and metacognition within self-regulated learning. In D. Schunk, & J. Greene (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 36–48). New York, NY: Routledge.
Winne, P., & Azevedo, R. (2014). Metacognition. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 63–87). Cambridge: Cambridge University press.
Winne, P., & Hadwin, A. (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., & Hadwin, A. (2008). The weave of motivation and self-regulated learning. In D. Schunk, & B. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297–314). New York: Taylor & Francis.
Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments, 7*(3), 225–240.
Wouters, P., van Nimwegen, C., van 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*(2), 249–265.