



Quantifying Scientific Thinking Using Multichannel Data With Crystal Island: Implications for Individualized Game-Learning Analytics

Elizabeth B. Cloude*, Daryn A. Dever, Megan D. Wiedbusch and Roger Azevedo

Laboratory for the Study of Metacognition and Advanced Learning Technologies, Department of Learning Sciences and Educational Research, University of Central Florida, Orlando, FL, United States

OPEN ACCESS

Edited by:

Carrie Demmans Epp,
University of Alberta, Canada

Reviewed by:

Herre Van Oostendorp,
Utrecht University, Netherlands
April Lynne Zenisky,
University of Massachusetts Amherst,
United States

*Correspondence:

Elizabeth B. Cloude
elizabeth.cloude@knights.ucf.edu

Specialty section:

This article was submitted to
Assessment, Testing and Applied
Measurement,
a section of the journal
Frontiers in Education

Received: 14 June 2020

Accepted: 07 October 2020

Published: 02 November 2020

Citation:

Cloude EB, Dever DA, Wiedbusch MD
and Azevedo R (2020) Quantifying
Scientific Thinking Using Multichannel
Data With Crystal Island: Implications
for Individualized Game-Learning
Analytics. *Front. Educ.* 5:572546.
doi: 10.3389/feduc.2020.572546

Quantifying scientific thinking using multichannel data to individualize game-based learning remains a significant challenge for researchers and educators. Not only do empirical studies find that learners do not possess sufficient scientific-thinking skills to deal with the demands of the twenty-first century, but there is little agreement in how researchers should accurately and dynamically capture scientific thinking with game-based learning environments (GBLEs). Traditionally, in-game actions, collected through log files, are used to define if, when, and for how long learners think scientifically about solving complex problems with GBLEs. But can in-game actions distinguish between learners who are thinking scientifically while solving problems vs. those who are not? We argue that collecting multiple channels of data identifies if, when, and for how long learners think scientifically during game-based learning compared to only in-game actions. In this study, we examined relationships between 68 undergraduates' pre-test scores (i.e., prior knowledge), degree of agency, eye movements, and in-game actions related to scientific-thinking actions during game-based learning, and performance outcomes after learning about microbiology with Crystal Island. Results showed significant predictive relationships between eye movements, prior knowledge, degree of agency, and in-game actions related to scientific thinking, suggesting that combining these data channels has the potential to capture when learners engage in scientific thinking and its relation to performance with GBLEs. Our findings provide implications for using multichannel data, e.g., eye-gaze and in-game actions, to capture scientific thinking and inform game-learning analytics to guide instructional decision making and enhance our understanding of scientific thinking within GBLEs. We discuss GBLEs designed to guide individualized and adaptive game-analytics using learners' multichannel data to optimize scientific thinking and performance.

Keywords: scientific thinking, multichannel data, game-based learning, game-based analytics, individualized game-based learning environments

1. INTRODUCTION

Scientific thinking has steered the crusade of discovery and changed the way in which we understand, interact, and exist in the world (Klahr et al., 2019, p. 67). As our communities are faced with persisting, and sometimes novel, socioeconomic, environmental, and health-related challenges (e.g., interrupting traditional classroom instruction and resorting to remote learning due to the COVID-19 pandemic), it is essential to equip future generations of learners with scientific-thinking skills (National Research Council, 2012; NASEM, 2018; Kuhl et al., 2019). Recent advances in learning technologies, such as game-based learning environments (GBLEs), have guided design principles (e.g., role of agency in fostering cognitive engagement underlying scientific thinking) that enhance learners' skills and knowledge related to scientific thinking and offer solutions to problems that learners face in formal education settings (e.g., learners being tasked with memorizing factual information outside of real-world application; Council et al., 2011; Deater-Deckard et al., 2013, p. 21; Morris et al., 2013, p. 607; Plass et al., 2020b). After decades of studying cognitive processes underlying scientific thinking for training and educational purposes (Dunbar and Klahr, 2012, chapter 34; Byrnes and Dunbar, 2014, p. 477), it has become clear scientific thinking relies on factors that are personal to each learner (Klahr and Dunbar, 1988, p. 1; Dunbar and Klahr, 2012, chapter 34). For example, prior knowledge plays a key role in learners' ability to formulate hypotheses. However, studies with GBLEs use an approach that generalizes across learners, failing to account for individual characteristics that may impact learners' ability to think scientifically. These issues further compound current problems associated with providing individualized instructions in GBLEs. If we do not know how to capture if, when, and for how long learners engage in scientific thinking with GBLEs, how do we provide data-driven, individualized instruction to meet learners' needs? Researchers face additional challenges when capturing scientific thinking with GBLEs because most studies rely on in-game actions, measured solely through log files, to define if, when, and for how long a learner is thinking scientifically about solving problems, such as the amount of time using a scanner to test evidence (Smith et al., 2019, p. 52). But we argue in-game actions do not provide enough information to identify whether learners are thinking scientifically about solving problems with GBLEs. For example, can researchers distinguish between a learner aimlessly engaging in an action (e.g., testing random food items) vs. a learner engaging in an action because they are thinking scientifically about solving a problem (e.g., testing food items based on current hypotheses developed from information gathered)? Other data channels such as eye movements and pre-test scores could supplement information captured via in-game actions to inform if, when, and for how long learners are thinking scientifically during game-based learning. In this study, we investigated whether eye movements, pre-test scores, degree of agency, in-game actions, and post-test scores identified if, when, and for how long learners were thinking scientifically about solving problems with GBLEs. Our findings provide implications for designing GBLEs to guide individualized and adaptive interventions based

on learners' individual needs using their multichannel data to optimize scientific-thinking skills and performance.

1.1. What Is Scientific Thinking?

Scientific thinking defines two types of thinking (Dunbar and Klahr, 2012, chapter 34; Klahr et al., 2019). The first is literally thinking about the content of science-related topics, such as the characteristics of a virus. The second type of thinking encompasses a set of reasoning strategies or cognitive processes, such as inductive and deductive reasoning, problem solving, as well as hypothesis formulation and testing (Dunbar and Klahr, 2012, chapter 34; Zimmerman and Croker, 2014, p. 245). These two types of scientific thinking have a codependent relationship with each other (Dunbar and Klahr, 2012, chapter 34; Klahr et al., 2019). For instance, a learner must think about the idiosyncrasies of a virus in relation to other infectious diseases to conceptually understand the content, which lays the foundation for reasoning about how a virus might spread in a research camp compared to bacteria. Because of this, learners' prior knowledge plays a crucial role in scientific thinking as related to reasoning strategies and other cognitive processes (Zimmerman and Croker, 2014, p. 245). However, few studies account for the role of learners' prior knowledge in their ability to think scientifically during problem solving with GBLEs. Klahr and Dunbar (1988) developed a framework called scientific discovery as dual search (SDDS) which accounts for prior knowledge by describing hypothesis formulation as relying on long-term memory to identify gaps in knowledge to drive information-gathering actions (Dunbar and Klahr, 2012, chapter 34; Zimmerman and Croker, 2014, p. 245). Because of this, SDDS has the potential to explain learners' individual characteristics which may influence their ability to think scientifically.

1.1.1. Scientific Discovery as Dual Search

Klahr and Dunbar (1988) conceptualize scientific thinking as searching within and between two problem spaces—the hypothesis space and the experimental space. Each space is distinguished by its own set of operations and representations. In the hypothesis space, learners search their long-term memory (i.e., prior knowledge) and/or metacognitive knowledge and experiences to formulate hypotheses, while the experimental space is guided by planning and investigating current hypotheses. The learner must search through their hypothesis testing results to inform the generation of alternative hypotheses in this space, or make conclusions about the state of the phenomenon in which they are studying. The model also suggests that in order to initiate successful scientific thinking, learners must engage in three key components. First, learners must gather information before they can formulate a hypothesis. This requires searching through memory and the environment in which the learner is learning. Second, learners formulate a hypothesis for testing based on their understanding. Last, once a hypothesis is formed, the learner tests their current hypothesis and evaluates the results. Based on the results, they can either come to a decision about the phenomena they are studying (i.e., reject the hypothesis), or use the results to formulate alternative testable hypotheses (i.e., accept the hypothesis; Klahr and Dunbar, 1988, p. 1). As such, the

amount of time spent gathering information is contingent on the learners' prior knowledge about the content, where the more time that learners spend gathering information, or searching through memory, the less time they have to allocate to other components of scientific thinking such as formulating and testing hypotheses. Previous research using SDDS to study scientific thinking found that a variety of individual characteristics contribute to scientific thinking and thus impact subsequent learning and performance outcomes (Lazonder and Harmsen, 2016), such as prior knowledge about the domain (Dunbar and Klahr, 2012, chapter 34) and types of scaffolds used to guide learners (Lazonder et al., 2010, p. 511; Mulder et al., 2011, p. 614; Taub et al., 2020b, p. 1). However, major gaps persist as this work has not been extended to problem solving with GBLEs.

1.2. Game-Based Learning Environments

GBLEs designed to foster scientific-thinking and problem-solving skills use a range of game features that influence learners' personally, such as enhancing their emotional engagement, perceived agency, and interest in learning (Plass et al., 2015, p. 258; Plass et al., 2020a, chapter 1). A number of meta-analyses provide evidence that the design of GBLEs amplifies knowledge and skill acquisition (Wouters et al., 2013, p. 249; Mayer, 2014; Clark et al., 2016, p. 79) by (1) providing an incentive structure such as granting rewards for completing tasks, (2) visual and auditory aesthetics used to engage learners, and (3) a narrative with clearly-defined rules that limit agency by requiring learners to finish tasks to successfully complete the objective of the GBLE [e.g., talking to a non-player character (NPC) to gather clues to move forward with other tasks]. However, these meta-analyses have not examined the role or impact of scientific thinking with GBLEs, nor have they provided actionable multichannel data collected during game-based learning to prescribe the most important features that can be used to individualize scientific thinking with GBLEs. Agency, or learners' ability to control their actions during learning (Bandura, 2001, p. 1), is a critical design within GBLEs and has been shown to impact scientific thinking (Taub et al., 2020b, p. 1). For instance, research has found that while affording total agency—i.e., no restrictions to learners' actions, results in increased engagement, motivation (Mayer, 2014; Shute et al., 2019, p. 59), and performance outcomes with GBLEs (Loderer et al., 2020; Plass et al., 2020b), there are mixed findings regarding agency and its impact on scientific thinking. GBLEs that provide total agency require learners to effectively engage in scientific thinking all on their own, where learners may become distracted by other activities or seductive but extraneous features that are unrelated to scientific thinking, such as exploring the environment or trying to game the system. Because of this, varying degrees of agency (e.g., partial vs. total agency) serve as an implicit scaffolding technique for developing scientific-thinking skills with GBLEs.

1.2.1. Agency That Supports Scientific Reasoning as Inconspicuous Scaffolding

Implicit scaffolds, such as partial agency, support knowledge, and skill acquisition by quietly changing the way learners interact with information, tools, and game elements built into GBLEs. For

example, Crystal Island is a narrative-centered GBLE designed to teach learners about microbiology using varying degrees of agency (Taub et al., 2020b, p. 1). The total agency condition affords learners absolute control over their actions during game-based learning, while the partial agency condition requires learners to engage in a predefined and fixed sequence of actions (e.g., learners must read all books within each building before testing hypotheses) with the assumption that the required in-game actions are beneficial for effective scientific thinking (e.g., gathering information prior to generating hypotheses), learning, and performance outcomes. Learners in the partial agency condition, however, are still given some control by allowing them to use tools at any point during game-based learning, such as recording clues and formulating hypotheses using the scientific worksheet. The no agency condition does not afford learners any control over their actions as they watched a scientific-thinking expert in 3rd person complete the game. A study investigating the effect of agency on scientific thinking and performance using Crystal Island found that learners in the partial agency condition demonstrated more scientific thinking in-game actions and higher performance outcomes relative to learners in the total or no agency conditions (Taub et al., 2020b, p. 1). However, because scientific thinking was solely measured using in-game actions, challenges exist as the authors cannot ensure learners were actually thinking scientifically vs. completing actions to move forward in the game. Other sources of data are critical to inform if, when, and for how long learners engage in scientific thinking during game-based learning.

1.3. Quantifying Scientific Thinking With GBLEs

A study by Shute et al. (2016) captured scientific thinking using the number of in-game actions learners initiated during problem solving and found that in-game actions were associated with successful problem solving and a strong predictor of performance compared to other methodological approaches such as problem-solving ability assessments (e.g., Raven's Progressive Matrices; Raven, 1941, p. 137; Raven, 2000, p. 1). Yet, major methodological and analytical limitations still exist in this work. Leveraging in-game actions that suggest scientific thinking does not capture information on whether the learner was actually thinking scientifically about solving the problems. For example, scientific thinking was operationally defined as learners' in-game actions that suggested analyzing resources. But can in-game actions capture when learners analyzed resources without information on where they were visually attending? Perhaps the learner was aimlessly playing the game and interacted with a resource by chance. A similar study by Taub et al. (2018) examined the sequence of in-game actions to quantify successful scientific thinking, where in-game actions were defined based on the relevance of learners' actions toward solving the problem during game-based learning (e.g., testing food items and pathogens relevant to the problem solution). Their results revealed that learners were more effective in their scientific thinking if their in-game actions suggested they tested fewer hypotheses that were more relevant to the problem solution compared to learners

who tested more hypotheses that were less relevant to the problem solution (Taub et al., 2018, p. 93). However, the authors used in-game actions alone to define scientific thinking, making the assumption that all learners, regardless of prior knowledge and other individual characteristics, are scientifically thinking about solving the problem during game-based learning. Still, can one be certain that when learners tested fewer hypotheses more relevant to the problem solution, they did not accidentally select a relevant hypothesis when using in-game actions data? It is critical to investigate whether other sources of data might supplement information captured via in-game actions suggesting scientific reasoning. Without capturing more information on what the learner is engaging with during game-based learning and accounting for individual characteristics such as prior knowledge, major gaps persist in research studying scientific thinking during problem solving with GBLEs.

1.4. Eye-Tracking Methodology

Including other data channels to supplement in-game actions during game-based learning may better inform when learners are actively engaging in scientific thinking. Decades of literature suggest eye-gaze data could be a promising direction in revealing implicit cognitive processing like scientific thinking (Scheiter and Eitel, 2017, p. 143). Adopting a cognitive psychology perspective, where learners visually attend is based on the size of the retina. The retina determines if and how much visual information is available for processing and discloses the learners' foci of attention and thus what information is being processed (van Zoest et al., 2017, p. 1555). A number of studies provide empirical evidence supporting this perspective, such that where participants look is indicative of their reasoning behaviors such as scientific thinking (Plummer et al., 2017, p. 1426; Miller Singley and Bunge, 2018, p. 445). For instance, Vendetti et al. (2017) showed that participants' saccades and fixations data were indicative of reasoning. Specifically, they found that participants' eye movements revealed optimal strategies for solving visual analogy problems when they were faced with distracting stimuli. Their findings showed that when participants attended to relevant relationships in stimuli, they were more likely to solve the analogy problems compared to participants who attended to distracting stimuli (Vendetti et al., 2017, p. 932). Several studies have also found that where learners visually attend is related to their intention such that the eye-gaze data predicted future in-game actions (Hillaire et al., 2009, p. 43; Rayner, 2009, p. 1457; Huang et al., 2015, p. 1049; Park et al., 2016, p. 796; Rajendran et al., 2018, p. 455) and reasoning behaviors (Bondareva et al., 2013, p. 229). A study by Munoz et al. (2011) analyzed eye-gaze data using an artificial neural network to assess whether these data predicted future actions. Their model achieved an accuracy rate above 83% for predicting future actions based on eye-gaze positions on the screen, such that the more often learners fixated on game elements within the game, the more likely they were to initiate in-game actions using those game elements in the future (Munoz et al., 2011, p. 47). A similar study by Huang et al. (2015) examined whether eye-gaze data predicted future actions. Their model predicted actions 1.80 s before the learner initiated the action based on

where they had previously fixated at an accuracy rate of 76% (Huang et al., 2015, p. 1049). As such, do these findings transfer to game elements in GBLEs which are intentionally designed to foster scientific thinking? Specifically, can eye-gaze data inform whether a learner is engaging in scientific thinking based on relationships between eye-gaze data and in-game actions related to scientific thinking with GBLEs? A study by Singh et al. (2018) investigated whether there were differences between a model with in-game actions vs. a model with both eye-gaze data and in-game actions collected during a multiplayer game in predicting future in-game actions. Their results found that the model with both eye-gaze and in-game actions data achieved a better fit with 71% accuracy for predicting future actions compared to the model with only in-game actions at 41% accuracy. The aforementioned studies demonstrate how eye-gaze data might inform scientific thinking in-game actions (Singh et al., 2018, p. 488). In sum, it is essential to use more than one data channel to capture and analyze scientific thinking during game-based learning. Eye-gaze data that supplements in-game actions may inform if, when, and how long learners are thinking scientifically during game-based learning.

1.5. Supporting Game-Based Learning Using Multichannel Data

GBLEs allow researchers to capture rich data about complex learning constructs (Taub et al., 2020a, chapter 9). Game-learning analytics (GLA) are techniques developed for capturing, storing, analyzing, and detecting critical information about what, when, and how long learners' interact with game elements such as tools and other resources while solving problems during game-based learning. GLA afford opportunities to guide instructional prescriptions and decision making based on what learners are doing to optimize and transform their learning experience based on their multichannel data (Freire et al., 2016, p. 1; Lang et al., 2017). As such, GLA opens a door for capturing scientific thinking with GBLEs. Since theoretical models and previous studies suggest individual characteristics of learners play a crucial role in their ability to think scientifically such as the degree of agency (Taub et al., 2020b, p. 1) and prior knowledge about the content being studied during game-based learning (Klahr and Dunbar, 1988, p. 1; Dunbar and Klahr, 2012, chapter 34), it is crucial to capture and analyze GLA to advance the science of learning and guide instructional decision making. A number of studies have used GLA to gain insight into the role of individual characteristics to analyze its relation to learning and performance outcomes. Giannakos et al. (2019) used learners' multichannel data to capture skill acquisition while they played a pac-man game. Skill acquisition was defined by motor-movement adaptation and decision making, where learners were required to control four buttons on a keyboard over the course of three sessions that increased in difficulty. Five sources of data were collected: eye-gaze—i.e., pupil diameter, fixation, saccades, and events (e.g., number of fixations, number of saccades, etc.), keystrokes, EEG, facial expressions of emotions, and wristband data—i.e., heart rate, blood pressure, temperature, and electrodermal activity. Their findings showed that keystrokes

demonstrated the lowest amount of model accuracy in predicting skill acquisition at an error rate of 39%. But, by fusing multichannel data, they achieved a 6% error rate in predicting skill acquisition (Giannakos et al., 2019, p. 108). Similarly, Alonso-Fernández et al. (2019) captured learners' interaction data during game-based learning tasks to examine its relation to learning outcomes. By building predictive models using data-mining techniques, they found that in-game actions capturing how learners' interacted with game elements during game-based learning were 89.7% accurate in their predictive ability, leaving approximately 11% of error. However, similar to the aforementioned studies, they found that by combining both the learners' pre-test scores before game-based learning and their in-game actions during game-based learning, their predictive model demonstrated an accuracy rate of 92.6%, reducing its error rate to <8% (Alonso-Fernández et al., 2019, p. 301). Similarly, Sharma et al. (2020) captured and analyzed learners' in-game actions and physiological data—i.e., facial expressions of emotions, eye tracking, EEG, and wristband data to predict learners' effort based on patterns in their multichannel data while they were learning. Their results showed that Hidden Markov Models could detect learners' effort from their multichannel data generated during learning in a way that fostered individual instructional prescriptions in real-time that were informed by patterns in learning behaviors (Sharma et al., 2020, p. 480). In sum, these studies highlight how important it is to analyze multichannel data captured over a learning session to gain insight into skill and knowledge acquisition with GBLEs. GLA techniques are useful in tracking, modeling, and understanding what learning processes are initiated and how each contributes to an individual learner's conceptual understanding of the domain, or the degree at which a skill is acquired and mastered. The value of GLA is emphasized based on its ability to reveal behaviors which may be detrimental to a learner's performance that go beyond in-game actions captured via log files, offering opportunities to detect maladaptive behaviors and intervene during learning activities to redirect learners toward optimal learning, skill acquisition, and performance outcomes. For example, if a learner has less prior knowledge about game content, or is showing elevated time spent examining game elements, then it could signal the learner's scientific thinking skills. Further, this information could guide an instructional intervention tailored to the specific needs of the learner based on their individual characteristics, in addition to when, with what, how often and how long they are engaging with game elements related to the overall objective of the learning session. Yet, contrary to these promising directions, a number of studies continue to use in-game actions alone to capture scientific thinking with GBLEs (Shute et al., 2016, p. 106).

1.6. Current Study

In this study, we examined college students' scientific thinking with GBLEs to determine optimal features for personalizing instruction during science learning. Current studies apply a generalized methodological and analytical approach by only accounting for performance and in-game actions logs to define if, when, and how long learners think scientifically about solving problems with GBLEs (Shute et al., 2016, p. 106; Smith et al.,

2019, p. 52; Taub et al., 2018, p. 93; Taub et al., 2020b, p. 1), ignoring individual characteristics of learners that have been shown to impact their ability to think scientifically during game-based learning. To address the challenges mentioned above, we investigated whether learners' multichannel data were related to scientific thinking during game-based learning and performance outcomes. Specifically, the objective of our study was to examine whether relationships existed between multichannel data—i.e., eye gaze, in-game actions, degree of agency, and pre-test and post-test scores. Our research questions were grounded in SDDS theory (Klahr and Dunbar, 1988, p. 1) and previous empirical evidence related to modeling scientific thinking to examine whether the proportion of time fixating on game elements related to scientific reasoning, pre-test scores, and degree of agency predicted the proportion of time interacting with game elements related to scientific reasoning and post-test scores. Our research questions and hypotheses are provided below:

Research Question 1: To what extent does the time fixating and interacting with scientific reasoning related game elements predict post-test scores after game-based learning, while controlling for prior knowledge and degree of agency? We hypothesize there will be predictive relationships between the time interacting with scientific reasoning related game elements and post-test scores after game-based learning, while controlling for pre-test scores and degree of agency. Our hypothesis is both grounded in Klahr and Dunbar (1988) SDDS theory and previous empirical evidence that multichannel data generated over a learning session [e.g., (Alonso-Fernández et al., 2019, p. 301), prior knowledge (Dunbar and Klahr, 2012, chapter 34), and the degree of agency (Lazonder et al., 2010, p. 511; Mulder et al., 2011, p. 614; Taub et al., 2020b, p. 1)] has been related to performance outcomes.

Research Question 2: To what extent does time fixating on scientific reasoning related game elements predict time interacting with scientific reasoning related game elements during game-based learning, while controlling for pre-test scores and degree of agency? We hypothesize there will be predictive relationships between the time fixating on scientific reasoning related game elements and the time interacting with scientific reasoning related game elements during game-based learning, while controlling for pre-test scores and degree of agency. Our hypothesis is based on empirical evidence suggesting that eye movements are related to learners' cognitive processing and future in-game actions (Hillaire et al., 2009, p. 43; Huang et al., 2015, p. 1049; Park et al., 2016, p. 796; Singh et al., 2018, p. 488).

Research Question 3: To what extent does the time fixating on non-scientific reasoning related game elements predict the time interacting with non-scientific reasoning related game elements during game-based learning, while controlling for pre-test scores and degree of agency? We hypothesize there will be predictive relationships between the time fixating on non-scientific reasoning related game elements and time interacting with non-scientific reasoning related game elements unrelated to scientific reasoning during game-based

learning, while controlling for pre-test scores and degree of agency. Our hypothesis is based on empirical evidence suggesting that eye movements are related to learners' cognitive processing and future in-game actions (Hillaire et al., 2009, p. 43; Huang et al., 2015, p. 1049; Park et al., 2016, p. 796; Singh et al., 2018, p. 488).

2. MATERIALS AND METHODS

2.1. Participants and Materials

A total of 138 learners were recruited from three large, public universities in North America. For this paper, a subset of 68 participants ($n = 68$; 68% female, age: $M = 20.01$, $SD = 1.56$) were included in our analyses because they met the inclusion criteria: complete eye-tracking, log file, and performance data and were randomly assigned to either the total agency ($n = 41$) or partial agency conditions ($n = 27$). Of the subsample, the majority identified as Caucasian (75%), while the remaining identified as Asian, Black, Hispanic or Latino, and Other. Additionally, the majority of the subsample reported rarely playing video games (37%), while the remaining reported occasionally (24%) or frequently (16%), or very frequently (9%) playing. Most participants reported an average level of video game playing skill (41%), whereas others reported none at all (13%), limited skills (21%), skilled or very skilled (25%). Sixty-three percent of participants reported playing video games 0–2 h per week (63%), while the remaining reported playing 3–4 h (16%), 5–10 h (7%), 10–12 h (12%), or more than 20 h (1%) per week. This study was approved by the Institutional Review Board prior to recruitment and informed written consent was obtained prior to data collection. To measure knowledge about microbiology, a 21-item, 4-option multiple choice assessment was administered before and after participants finished the game, regardless of whether they had solved the mysterious illness plaguing the island. Participants answered between 6 and 18 correct items ($Med = 12$, $M = 57\%$, $SD = 0.13$)¹ on the pretest, and between 10 and 19 correct items ($Med = 15$, $M = 72\%$, $SD = 0.12$) on the post test. The assessments contained 12 factual (e.g., “What is the smallest type of living organism?”) and nine procedural questions (e.g., “What is the difference between bacterial and viral reproduction?”). Several self-report questionnaires were also administered to participants before and after game-based learning to gauge their emotions, motivation, self-efficacy, and cognitive load². Scientific thinking was measured using a combination of log files and eye-gaze behaviors (see Coding and Scoring subsection for details). Game play duration ranged from 69 to 94 min ($M = 83$, $SD = 17.5$).

2.2. Experimental Design

Crystal Island was designed with three experimental conditions: (1) total agency where participants had the autonomy to make

their own decisions during game-based learning without any restrictions, (2) partial agency where participants had limited autonomy to make their own decisions during game-based learning due to restrictions around the sequence of buildings they could go to and tools they could use, and (3) no agency conditions where participants watched in third-person as another player worked toward identifying the mysterious pathogen in 3rd person. Upon beginning the game, participants within the total agency condition could go to any of the buildings and open whichever tool they deemed important toward identifying the unknown pathogen at any time they chose, while participants in the partial agency condition were required to visit the nurse first and discuss the symptoms the inhabitants were experiencing before being required to follow a “golden path” of steps to solve the mystery. The partial agency condition was designed to optimize scientific-reasoning actions by requiring participants to first gather information related to the unknown pathogen (e.g., symptoms), read specific books and research articles touching on types of pathogens, and then experimentally test their generated hypotheses. The no agency condition was designed as a means to model what scientific-reasoning actions should look like during game-based learning with Crystal Island.

2.3. Crystal Island

Crystal Island is a GBLE designed to foster and improve scientific-reasoning skills through an objective-based quest during complex problem solving. The science content in the game aligns with the Standard Course of Study Essential Standards for Eighth-Grade Microbiology (McQuiggan et al., 2008, p. 510). The problem-solving activities emphasize the practice of scientific inquiry as called for by the Next Generation Science Standards, and they also align with standards from the Common Core State Standards for English Language Arts on Reading: Informational Text. Participants were tasked with identifying a mysterious pathogen infecting inhabitants in a research camp on an isolated island (Figure 1, adopting the role of a Center for Disease Control and Prevention agent to identify an unknown pathogen that infected a team of researchers (Rowe et al., 2011, p. 115). Participants were instructed to provide an accurate diagnosis and treatment solution by testing possible transmission sources (e.g., food items such as eggs and cheese), talking to NPCs on the island about their expertise or symptoms, and collecting information through resources (e.g., books and research articles) scattered around the island in order to complete the game. During the game, participants moved through a variety of buildings, each containing different game elements providing information related to pathology including NPCs, research articles, books, posters, food items, and devices to test hypotheses that foster scientific thinking. Specifically, Crystal Island was built with a (1) dining hall (where participants can collect food items and interact with the cook—i.e., NPC to assess which food has been served recently; see Figure 2) where an orange is illustrated as a food item); (2) laboratory (where participants can test food items picked up throughout the game to assess whether they have been infected with a pathogen; see Figures 1, 3; (3) infirmary (where participants can

¹The mean was calculated using a ratio score of total correct items over total items on the pre/post-test assessments. Median was reported for the total number of correct items.

²We do not provide information about the self-report questionnaires because they were not included in the analyses. Readers are encouraged to email the corresponding author to request information about the scales administered.



FIGURE 1 | Crystal Island environment. **(Top left)** Living quarters; **(Top right)** Laboratory; **(Bottom left)** Island; **(Bottom right)** Dormitory.



FIGURE 2 | Generating hypotheses related game elements. **(Left)** Diagnostic worksheet. **(Right)** Dining hall where participants can select food items; circled game elements = Areas of Interest for a food item and backpack (how participants access the diagnostic worksheet).

interact with a nurse and sick patients who provide information about their symptoms; see **Figures 1, 4**); and (4) dormitory (**Figure 1**) as well as (5) living quarters (where participants can interact with other team members in the research camp; **Figure 1**). Using the information gathered from the various island locations, participants can record clues and hypotheses for later testing using a range of tools provided during learning with Crystal Island.

2.3.1. Tools Designed to Foster Scientific Thinking

Six tools were built into Crystal Island to foster scientific thinking and help participants navigate the game challenges and mystery. First, *books*, *posters*, and *research articles* (see **Figure 1**) were scattered around the island and served as some of the main sources of information about microbiology and pathology. The books, articles, and posters were originally written by a former middle-school science teacher and ranged in topic,



FIGURE 3 | Experimental testing related game elements. **(Left)** Concept matrix content and feedback (displayed in red when participant has incorrect answers). **(Right)** Scanner for testing food items in laboratory.

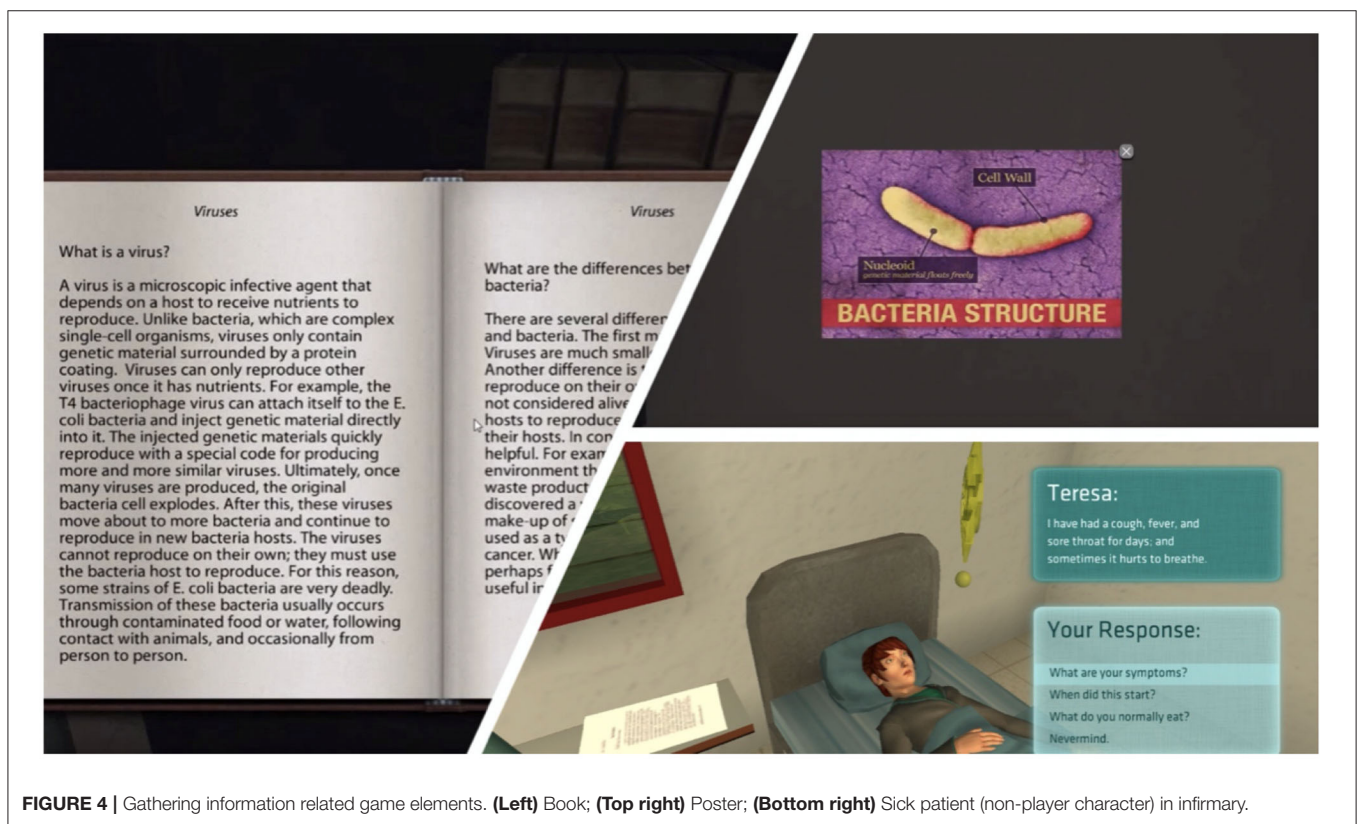


FIGURE 4 | Gathering information related game elements. **(Left)** Book; **(Top right)** Poster; **(Bottom right)** Sick patient (non-player character) in infirmary.

length, and text difficulty. Each of the sources explained distinct characteristics related to certain types of pathogens and the ways to treat (1) bacteria, (2) viruses, (3) parasites, (4) fungi, and (5) genetic diseases. Pathogens covered in books, research articles and posters included influenza, tapeworm, anthrax, salmonella, *E. coli*, polio, ebola, and sickle-cell anemia. Participants within the partial agency condition were required to “read” every book, poster and research article available to them. However, it is important to note that the system counted a book, poster, and research article as read if it was opened (for however brief). While log files also indicated how long the book remained

open, participants could have not actually chosen to read, but instead looked elsewhere on the screen. Participants within the total agency condition, however, were not required to read any of the books, posters, or research articles if they chose not to. Within books and research articles, participants had an opportunity to test their understanding of the new information they had just read about using a *concept matrix* (see **Figure 3**). Concept matrices were presented as a matrix where participants matched pathogen types (e.g., virus) to their associated and distinct characteristics (e.g., reproduce quickly in living host cells). Participants were given a total of three attempts to provide

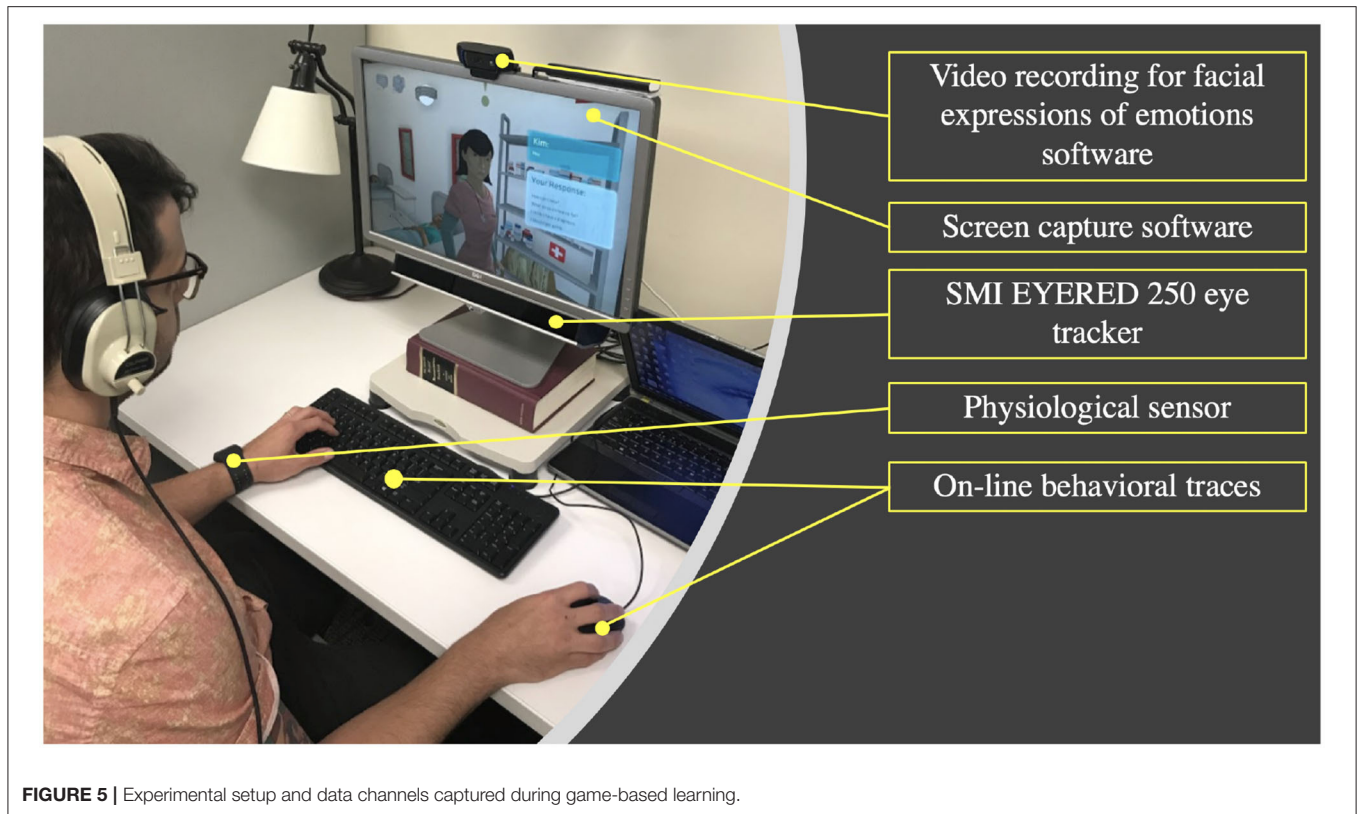


FIGURE 5 | Experimental setup and data channels captured during game-based learning.

correct information in the mapping exercise. If participants failed to answer correctly within the first 3 trials, the correct answers were displayed to them (see **Figure 3**). Participants within the partial agency condition were required to successfully complete (or at least attempt three times) every concept matrix. This was done with hopes participants would be more likely to read all of the information available within the game and therefore increase their microbiology knowledge and make more informed deductions and hypotheses regarding the mystery. However, given that the correct answer was provided to participants after three failed attempts, participants could be incentivized to game the system for non-relevant (or uninteresting) books and research articles. Because participants within the total agency condition were not required to complete concept matrices if they chose not to, it is possible they were less likely to game the system as there was no reward for doing so or punishment for not completing them. Participants also had access to a *scanner* in the laboratory building (see **Figure 3**), where they could test food items they hypothesized as transmitting the pathogen to the sick researchers. During the game, participants were also provided with a *diagnosis worksheet* (see **Figure 2**) that allowed them to record information related to (1) different types of pathogens and their distinct characteristics; (2) symptoms the sick inhabitants were experiencing; and (3) hypotheses about the transmission source of the pathogen (e.g., orange, milk, eggs). For example, participants could keep track of food items they had previously tested to foster their scientific thinking (e.g., if cheese had negative results for bacteria, participants could deduce

either the cheese was not the source of the illness or bacteria was not the pathogen). Items did not have to be scanned in order to correctly solve the mystery, however it was available to test hypotheses generated from other clues. Participants within the partial agency condition were required to scan items, while those in the total agency condition could choose if they wanted to. While participants could fill out and edit the diagnosis worksheet at any time during the game, they had to submit an accurate diagnosis, transmission source, and treatment solution to successfully solve the mystery and complete Crystal Island.

2.4. Procedure

When participants arrived at the laboratory, a researcher confirmed their identity and ensured no clothing, hair, or glasses (including eye conditions such as astigmatism) would interfere with the eye-tracker calibration and data collection on eye movements. First, participants completed informed, written consent and then were instrumented with an electro-dermal bracelet and calibrated to the eye tracker and facial expressions of emotions software (see **Figure 5** for experimental setup)³. After successful calibration, participants were instructed to complete several questionnaires gauging motivation, emotions, and self-efficacy as well as a 21-item, multiple choice pre-test assessment on microbiology. Next, instrumented participants

³For this study, we only analyzed data collected via the eye-tracker, performance assessments, and keyboard stroke/mouse clicks. As such, we do not provide information about other data channels, but readers are encouraged to email the corresponding author to request more information.

started learning with Crystal Island while we collected their multichannel data. On average, it took participants 81 min ($SD = 23$ min) to identify the unknown pathogen and a correct treatment solution. If participants did not solve the mysterious illness within 90 min, they were exited from the game. Immediately after, participants were instructed to complete several questionnaires gauging motivation, emotions, and presence in addition to a 21-item, multiple choice post-test assessment on microbiology. Upon completing the post-test session, participants were paid \$10/hr (up to \$30), debriefed, and thanked for their time and participation.

2.5. Apparatus

2.5.1. Eye-Gaze Behaviors

To record eye-gaze behaviors, an SMI EYERED 250 eye tracker (Center, 2014) was used in this study and detected pupil and fovea location using infrared light. We used a nine-point calibration and configured the eye tracker to capture eye-gaze data at a sampling rate of 30 Hz, capturing relatively small eye movements at an offset of <0.05 mm. Eye-gaze fixations were processed in iMotions software (iMotions, 2014), which provided granular *post-hoc* analysis for creating dynamic areas of interest (AOIs) around game elements related to scientific reasoning (e.g., time spent fixating on complex text was defined as information gathering; **Figure 4**). We defined two types of AOIs: (1) around game elements in which the learner was not interacting with (e.g., fixating on a book from a distance), and (2) around the content of the game elements in which the learner was interacting with (e.g., opening a book and fixating on the text). These two types of AOIs distinguished between the eye-gaze and interaction variables, where the AOIs capturing interaction content was used to define when learners were engaging with materials by combining the data with in-game actions.

2.5.2. In-game Actions

In-game actions were recorded and time-stamped in log files when participants used the mouse and/or keyboard for analyses. Specifically, this data channel provided event- and time-based actions over the course of game-based learning. Throughout this paper, we refer to this data channel as “interaction elements” since it signaled when participants interacted with game elements.

2.6. Coding and Scoring

2.6.1. Performance Measures

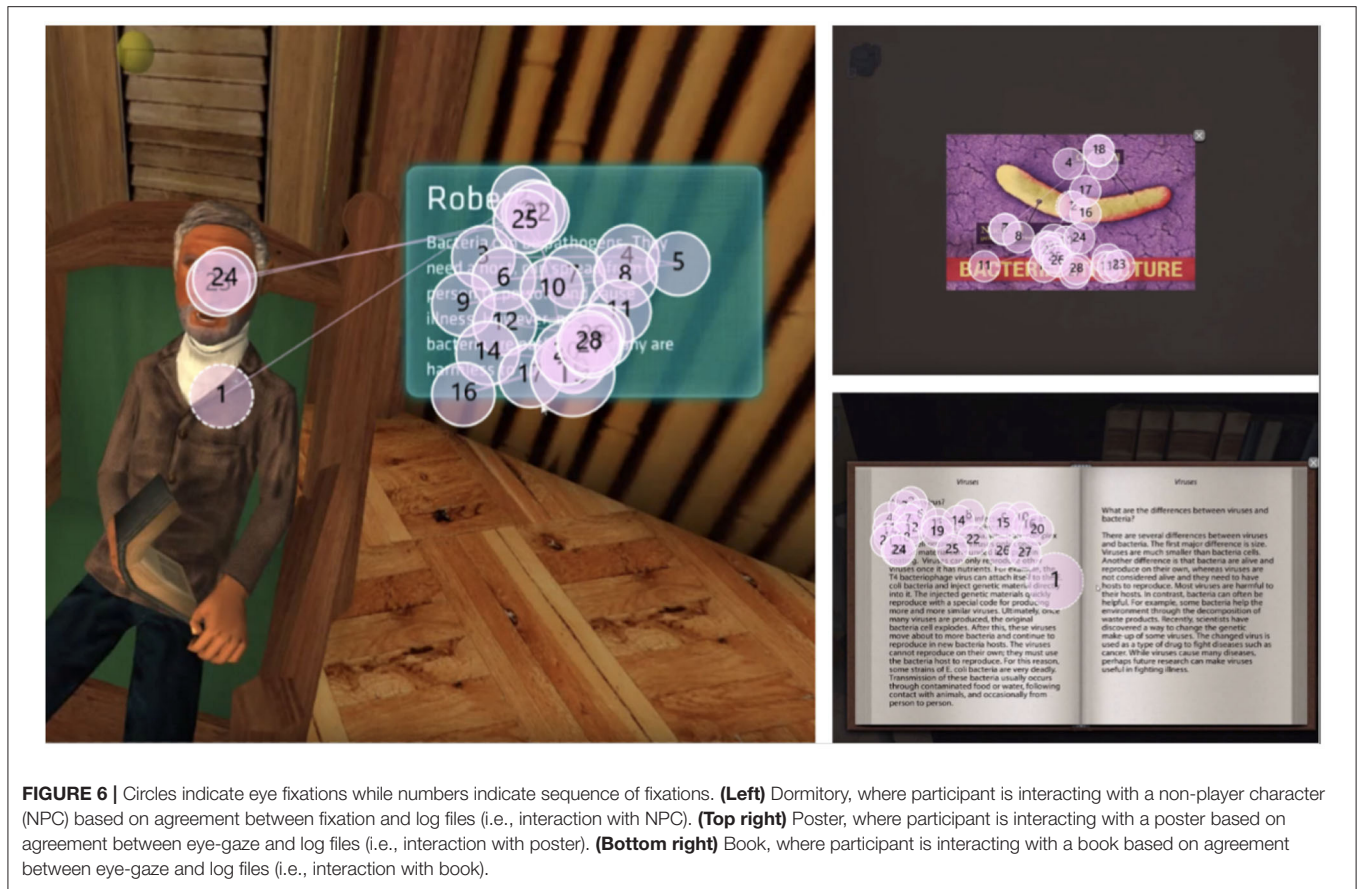
Prior knowledge and knowledge acquired during game-based learning was measured by creating a ratio of correct responses over total items on the pre- ($M = 0.57$, $SD = 0.13$) and post-test assessments ($M = 0.72$, $SD = 0.12$).

2.6.2. Fixations and Interactions With Game Elements

Game elements were categorized as either related to scientific reasoning (e.g., books, research articles, etc.) or non-scientific reasoning (e.g., doors, furniture, etc.) grounded within Klahr and Dunbar (1988) SDDS theory. Scientific reasoning game elements were further separated into three variables: (1) *gathering information*, (2) *generating hypotheses*, and (3) *experimentally testing hypotheses*. To define scientific reasoning based on

learners' interacting with game elements, we aligned in-game actions with eye-tracking data in order to measure two behaviors: (1) when a participant fixated on game elements, but did not interact with same game elements, and (2) when a participant fixated and interacted with game elements at the same time (i.e., interacting and fixating on game elements to ensure their engagement with game elements). The time a participant spent interacting⁴ with a game element required that the log file not only show that the game element was interacted with (e.g., a book opened, non-player character asked a question, or the diagnosis worksheet edited), but also that the participant was fixating on the content of the game element with which they were interacting with (see **Figure 6**). For example, if a participant fixated on a book on a table but did not touch it, this was considered a fixation. If they opened the book, but the eye-tracking data did not suggest they fixated on the book's content (e.g., looked elsewhere, opened and closed the book quickly, etc.), this was not considered interacting with a game element. However, if participants opened the book and fixated on the content, this was counted as the participant interacting with the book. Interactions were defined in this manner to ensure that the participant was engaging with the game element and so that we did not bias our analysis to consider interactions where the participant was not fixating on the content (i.e., accidentally interacting with a game element). Additionally, it helped prevent making the assumption that log files were indicative of scientific reasoning—that is, we do not have to assume that all in-game actions captured within the log files are meaningful. As such, the eye-tracking data helped augment and contextualize the log files to make a more meaningful interpretation and highlights the novelty of our approach compared to methods of only using log files. This is especially important for the partial condition participants who were required to complete certain actions before moving on. While the log file might alone indicate participants completed all actions in the partial agency condition, eye-tracking might illustrate other behaviors such as gaming the system. Additionally, for participants within the total agency condition, we would be able to determine if some actions were more exploratory search actions vs. scientific reasoning actions. Further, fixations were defined using gaze points within 1 degree visual angle exceeding at least 250 ms (Salvucci and Goldberg, 2000, p. 71) on AOIs described in section 2.5.1. Next, the fixations were aggregated across different element AOIs. Additionally, the time a participant spent fixating or interacting with game elements were aggregated across the categories described above [i.e., scientific reasoning elements (further grouped as either gathering information, generating hypothesis, or experimentally testing hypothesis) and non-scientific reasoning elements]. We then used these categories to find the proportion of time either fixating or interacting with elements compared to total time spent trying to solve the mysterious illness with Crystal Island. This was done to help control for the time differences produced by condition requirements of the partial agency group. It is

⁴Throughout this paper, we use interaction to signal that participants were not only interacting with game elements via in-game actions, but also fixating on game elements during their interaction with the game element via eye gaze.



important to note that the AOIs of game elements used to define fixations were different than the AOIs used to define interactions (see Tables 1, 2) which outline all of the various AOIs that were used to distinguish and categorize interactions and fixations with elements for our analysis). This is due to new visuals being shown when items were clicked on. For example, clicking on a book opened it up to content that was previously not visible. As such, fixating on a book was defined as a fixation on a closed book, while an interaction with a book was when a participant opened a book and fixated on the content in the book.

2.7. Statistical Analyses

We cleaned and processed our data in R (Version 3.6.2; R Core Team, 2013) using “read_xl” (Wickham and Bryan, 2017), “dplyr” (Wickham et al., 2018), and “reshape2” (Wickham et al., 2007, p. 1) packages for the data wrangling, manipulation, and melting features. Utilizing the “bestNormalize” package (Peterson and Cavanaugh, 2019, p. 1) and “plot” function from the “base” package (R Core Team, 2013), non-normally distributed variables were transformed into a normal distribution. Specifically, we used log, square root, and ordered quartile transformations, and in some cases, standardization. All variables were normalized or standardized with the exception of pre- and post-test ratio scores as these were normally distributed. Next, 12 participants were eliminated as they demonstrated

TABLE 1 | Fixations on game elements related to scientific and non-scientific reasoning.

Variables	Areas of interest on game elements	Operational definitions
Information gathering	All books, research articles, posters, and non-player characters	
Hypothesis generation	Food items, diagnosis worksheet, and backpack	
Experimental testing	Concept matrices and scanner	Time fixating on game elements for more than 250 ms without selecting, opening, editing, or viewing content of game elements/total time during game-based learning
Non-scientific reasoning	Doors, buildings, trees, plants, pots, settings, rocks, keys, menu icons, lights (e.g., bulbs), turbine, fire pit, treads, trophies, water, kitchen, appliances (e.g., refrigerator), lab equipment (e.g., flasks), windows, and furniture	

significant outlying observations via Grubb’s test (Grubbs, 1969, p. 21). Upon building the models, we used the “stepAIC” function from the “MASS” package to conduct stepwise model

selection using Akaike information criterion (AIC) for our first research question (Venables and Ripley, 2002, p. 271). The “summary” function from the “base” package was used to access object and model statistics, while the “ggplot2” package was used to visualize models and corresponding relationships among variables (Wickham, 2016). To visualize interaction terms, packages “tidyverse” (Wickham et al., 2019, p. 1686), “sjPlot” (Lüdtke, 2018b), and “sjmisc” (Lüdtke, 2018a, p. 754) were used. To probe the relation between interactions, we used the package, “interactions,” by calling the “probe_interaction” function (Long, 2019).

3. RESULTS

3.1. Preliminary Analyses

To examine the potential effect of the experimental manipulation on the results (i.e., level of agency), we conducted independent two-sample *t*-tests on all variables to assess if there were

differences between conditions. Analyses revealed no significant differences in both pre- and post-test scores between conditions ($ps > 0.05$) and so we did not include condition to maintain a parsimonious model for the first research question. For the second and third research questions, we found significant differences in time interacting with elements related to gathering information, $t_{(60)} = -3.09, p = 0.003$, and generating hypotheses, $t_{(50)} = 2.29, p = 0.026$, between experimental conditions and so we included condition in each equation ($ps < 0.05$; see Figure 7, Table 3). For research questions 2–3, we did not use the AIC method to select the best fit model because the predictor variables were empirically (i.e., literature suggests prior knowledge and agency impacts scientific thinking), theoretically (i.e., the SDDS model suggests prior knowledge impacts scientific thinking), and statistically justified (e.g., significant differences in eye-gaze and interaction data between conditions; see section 1.6 for more details).

TABLE 2 | Interactions with game elements related to scientific and non-scientific reasoning.

Variables	In-game actions with overlapping areas of interest	Operational definitions
Information gathering	All book content, research article content, poster content, and non-player character dialogue	
Hypothesis generation	Food items scanned and diagnosis worksheet	Time fixating on game elements for more than 250 ms while selecting, opening, editing, or viewing content of game elements/total time during game-based learning
Experimental testing	Concept matrix content, concept matrix feedback, and scanner	
Non-scientific reasoning	Movement to different locations during game-based learning	

3.2. To What Extent Does Time Fixating and Interacting With Scientific Reasoning Related Game Elements Predict Post-test Scores After Game-Based Learning, While Controlling for Pre-test Scores?

To examine whether relationships existed between proportion of time fixating and interacting with game elements related to scientific reasoning during game-based learning and post-test scores, while controlling for pre-test scores, we used AIC stepwise selection. Separate models for fixations and interactions with game elements were built, because including these data in the same equation to model performance created multicollinearity issues ($VIF > 10$). The first model was built using eye-gaze data to assess its relation to post-test scores while controlling for pre-test scores. A baseline model was determined using stepwise selection via AIC (Akaike, 1974, p. 716). The AIC method revealed that the best model was a

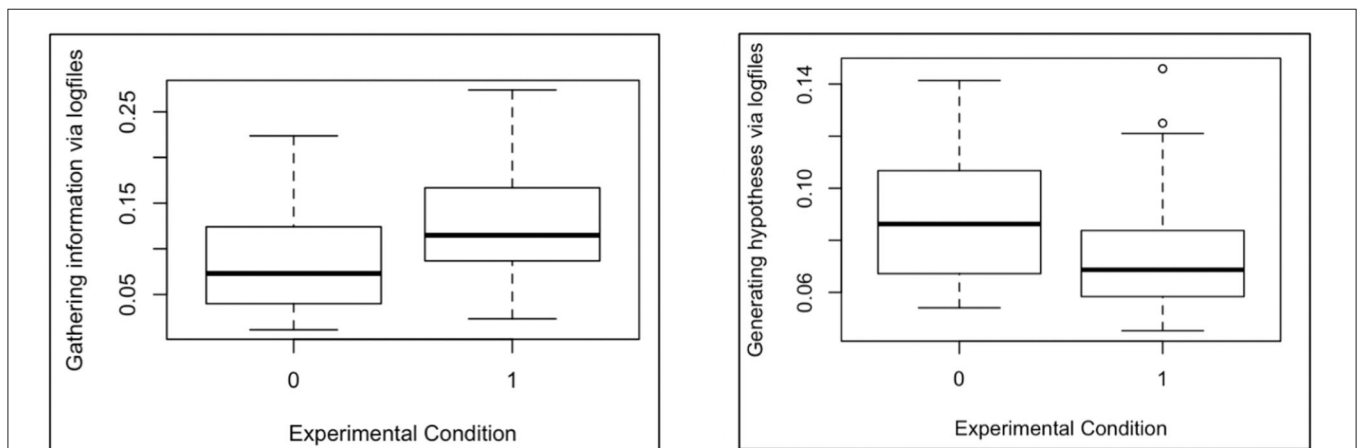


FIGURE 7 | Significant differences in proportion of time spent interacting with elements related to gathering information and generating hypotheses between experimental conditions; 0 = full agency, 1 = partial agency.

TABLE 3 | Descriptive statistics.

Variables	Condition	Mean (SD)	Median	1st Quartile	3rd Quartile
Information gather AOI	1	0.03 (0.03)	0.02	0.014	0.042
Information gather LOG	0	0.02 (0.02)	0.02	0.007	0.038
Hypothesis generate AOI	1	0.12 (0.06)	0.1	0.070	0.152
Hypothesis generate LOG	0	0.08 (0.05)	0.08	0.041	0.113
Experimental test AOI	1	0.04 (0.03)	0.02	0.012	0.044
Experimental test LOG	0	0.03 (0.03)	0.03	0.017	0.041
Pre-test scores	1	0.09 (0.03)	0.07	0.061	0.106
Post-test scores	0	0.09 (0.03)	0.09	0.068	0.111
Total game play	1	0.00 (0.00)	0.00	0.000	0.003
	0	0.00 (0.00)	0.00	0.000	0.001
	1	0.21 (0.09)	0.18	0.142	0.279
	0	0.22 (0.1)	0.19	0.145	0.261
	1	0.56 (0.14)	0.62	0.476	0.667
	0	0.55 (0.13)	0.52	0.48	0.62
	1	0.73 (0.12)	0.76	0.667	0.810
	0	0.70 (0.13)	0.71	0.62	0.76
	1	5586.15 (967.59)	5528	4616	6247
	0	4556.51 (895.45)	4345	3974	5098

Condition: 1 = partial agency, 0 = full agency.

simple linear regression equation, where pre-test scores were included as the only predictor variable (see **Table 4**). The fitted model estimated that average post-test scores increased by 0.43 points for each correct item on the pre-test assessment, where pre-test scores explained approximately 22% of the variance in post-test scores. As such, time spent fixating on game elements related to scientific and non-scientific reasoning during game-based learning were unrelated to post-test scores ($ps > 0.05$).

The second model was built using interaction with elements data to assess its relation to post-test scores, while controlling for pre-test scores. The AIC method indicated that the best model fit was a multiple linear regression equation, where pre-test scores and proportion of time interacting with game elements related to gathering information and generating hypotheses were included as predictor variables (see **Table 5**). The fitted model estimated that average post-test scores increased by 0.53 points for each correct answer on the pretest and increased by 0.24 points for each second increase based on a three-way interaction between proportion of time interacting with game elements related to gathering information and generating hypotheses, as well as pre-test scores correct item on the pre-test assessment. These predictors explained approximately 30% of the variance in post-test scores (see **Figure 8**). To probe the complexity of the three-way interaction relationship, we used a robust statistical technique known as the Johnson-Neyman interval (Johnson and Neyman, 1936, p. 57). The Johnson-Neyman technique relies on the range of values of the moderators (i.e., time spent interacting with elements

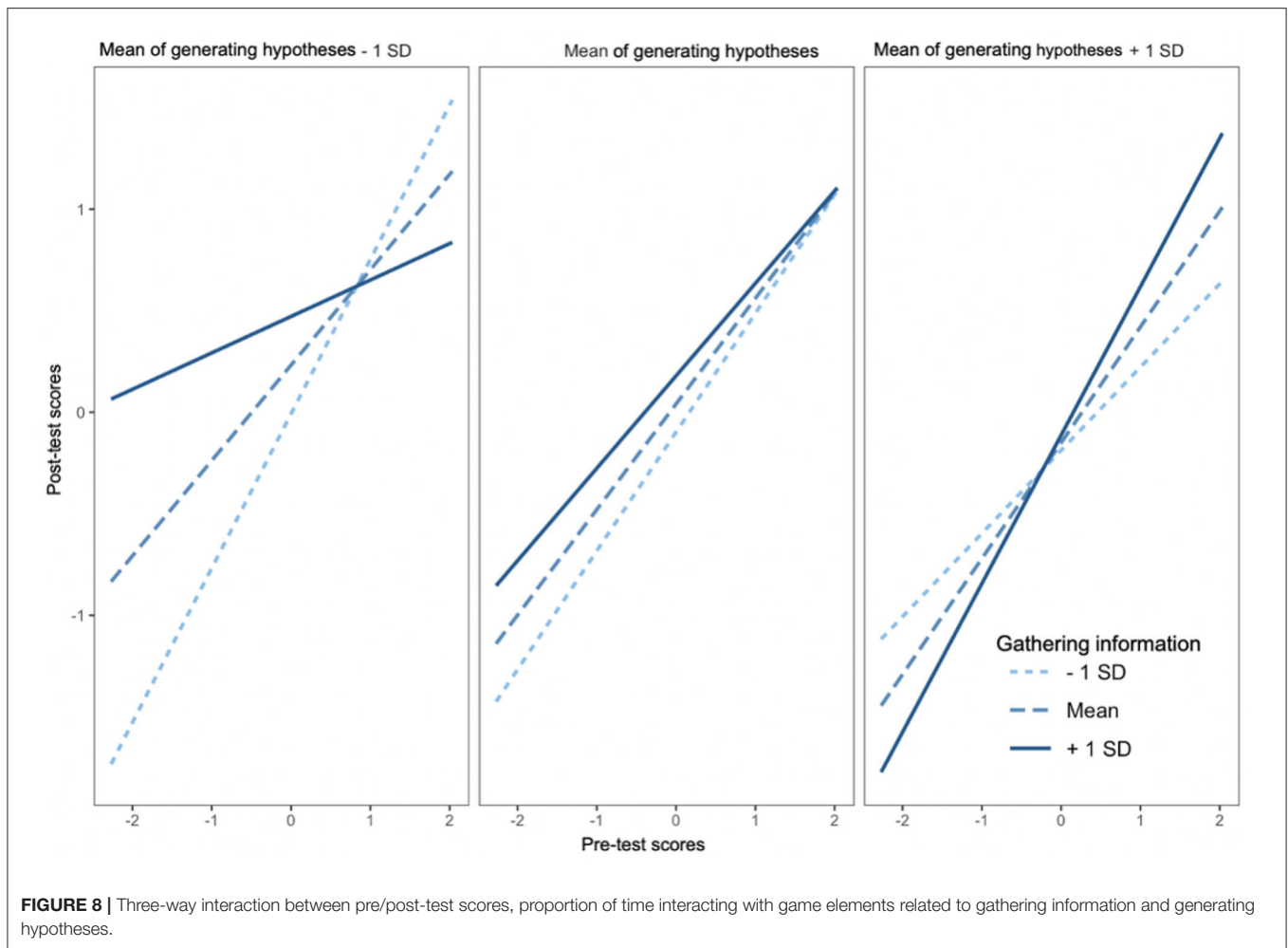
TABLE 4 | Modeling relationships with post-test scores; LOG, fixation, and interactions with game elements; AOI, fixations on game elements without interaction; * $p = 0.05$; *** $p < 0.05$.

Predictor variables	β	Standard error	t
Information gathering AOI	—	—	—
Hypothesis generation AOI	—	—	—
Experimental testing AOI	—	—	—
Non-scientific reasoning AOI	—	—	—
Condition	—	—	—
Pre-test scores	0.43***	0.10	4.48
Pre-test*Condition	—	—	—
Adjusted R^2		0.222	
F		20.08***	
Information gathering LOG	—	—	—
Hypothesis generation LOG	-0.21*	0.11	-1.90
Experimental testing LOG	—	—	—
Non-scientific reasoning LOG	—	—	—
Condition	—	—	—
Pre-test scores	0.53***	0.11	4.86
Pre-test*Condition	—	—	—
Pre-test*Information*Hypothesis	0.24***	0.12	2.03
Adjusted R^2		0.296	
F		5.024***	

TABLE 5 | Johnson-Neyman simple slope analysis modeling relationships with post-test scores.

Moderating variables	β	Standard error	t
PRE-TEST SCORES			
Hypothesis generation -1 SD			
Information gathering -1 SD	0.76***	0.25	2.99
Information gathering mean	0.47***	0.16	2.91
Information gathering +1 SD	0.18	0.26	0.70
Hypothesis generation mean			
Information gathering -1 SD	0.58***	0.16	3.77
Information gathering mean	0.52***	0.11	4.75
Information gathering +1 SD	0.46***	0.17	2.67
Hypothesis generation +1 SD			
Information gathering -1 SD	0.41***	0.13	3.04
Information gathering mean	0.57***	0.15	3.90
Information gathering +1 SD	0.73***	0.23	3.17

related to gathering information and generating hypotheses), where the slope of the predictor (i.e., pre-test scores) is significant compared to non-significant at an alpha level of 0.05. In our analysis, we considered pre-test scores as our focal predictor and proportion of time interacting with game elements



related to gathering information and generating hypotheses as moderators hypothesized to affect the relationship between pre- and post-test scores (see **Figure 8** for visualizations of pre- and post-test relationships as the moderators change from -1 standard deviation, to the mean, and $+1$ standard deviation). Refer to **Table 5** for a breakdown of how the relationships between post- and pre-test scores change based on the proportion of time interacting with game elements related to gathering information and generating hypotheses. In sum, the analysis suggested a significant positive relationship between pre- and post-test scores was present, except when participants spent <1 standard deviation ($SD = 0.025$) away from the mean ($M = 0.083$) generating hypotheses, but more than 1 standard deviation ($SD = 0.061$) away from the mean ($M = 0.102$) when gathering information. In other words, when participants spent a larger proportion of time during game-based learning gathering information compared to the relative average, and spent a smaller proportion of time generating hypotheses compared to the relative average, the relationship between pre- and post-test scores was not significant ($p > 0.05$).

3.3. To What Extent Does Time Fixating on Scientific Reasoning Related Game Elements Predict Time Interacting With Scientific Reasoning Related Game Elements During Game-Based Learning, While Controlling for Pre-test Scores and Degree of Agency?

To examine whether relationships existed between proportion of time fixating on game elements related to scientific reasoning and the proportion of time interacting with game elements related to scientific reasoning during game-based learning, while controlling for pre-test scores and experimental conditions, we built multiple linear regression equations. Specifically, we built separate models for each scientific-reasoning process: information gathering, hypothesis generation, and experimental testing, since including these variables in the same equation created major multicollinearity issues ($VIF > 10$). For each model, fixation variables were included as predictors while interaction with elements were included as targets. First, we built a model to examine relationships between the proportion of time

fixating on game elements related to gathering information and interacting with game elements related to gathering information, while controlling for pre-test scores and experimental conditions. A significant multiple linear regression equation was found, $F_{(7,60)} = 3.243, p < 0.001$, for the proportion of time interacting with game elements related to gathering information between total and partial agency conditions and a two-way interaction between pre-test scores and proportion of time fixating on game elements related to gathering information (see **Table 6** for beta coefficients and adjusted R^2). The fitted model estimated that the average proportion of time interacting with game elements related to gathering information decreased by 0.03 if participants were assigned to the total agency condition relative to the partial agency condition. We also found a significant two-way interaction, where the relationship between interactions and fixations for elements related to gathering information was affected by pre-test scores, such that the lower pre-test score (e.g., 44% correct items on pre-test) revealed a positive relationship between interactions and fixations for elements related to gathering information. However, the higher pre-test scores revealed a negative relationship between interactions and fixations for elements related to gathering information. Together, the predictors explained approximately 19% of the variance in the proportion of time interacting with game elements related to gathering information during game-based learning.

Our second model was built to examine relationships between the proportion of time fixating on game elements related to gathering information via interactions and fixations for elements related to generating hypotheses, while controlling for pre-test scores and experimental conditions. A significant multiple linear regression equation was found, $F_{(7,60)} = 2.225, p = 0.044$, suggesting relationships between interactions and fixations for elements related to generating hypotheses and experimental conditions (see **Table 6** for beta coefficients and adjusted R^2). The fitted model estimated that the average

proportion of time interacting with game elements related to generating hypotheses increased by 0.96 for each second increase in the proportion of time fixating on game elements related to generating hypotheses (see **Figure 9**), as well as the average proportion of time spent interacting with elements related to generating hypotheses decreased by 0.02 if participants were assignment to the partial agency condition compared to the total agency condition. Together, the predictors explained approximately 11% of the variance in the proportion of time interacting with game elements related to generating hypotheses during game-based learning. Finally, our last model was built to examine relationships between interactions and fixations for elements related to experimental testing, while controlling for pre-test scores and experimental conditions. We did not find a significant multiple linear regression equation, suggesting there were no relationships between interactions and fixations for elements related to generating hypotheses, while controlling for pre-test scores ($p > 0.05$). In sum, our analyses revealed a significant two-way interaction where pre-test scores moderated relationships between interactions and fixations for elements related to gathering information. We also found significant relationships between experimental conditions and interactions and fixations for elements related to generating hypotheses. We did not find relationships between interactions and fixations for elements related to experimental testing ($p > 0.05$).

3.4. To What Extent Does Time Fixating on Non-scientific Reasoning Related Game Elements Predict Time Interacting With Non-scientific Reasoning Related Game Elements During Game-Based Learning, While Controlling for Pre-test Scores and Degree of Agency?

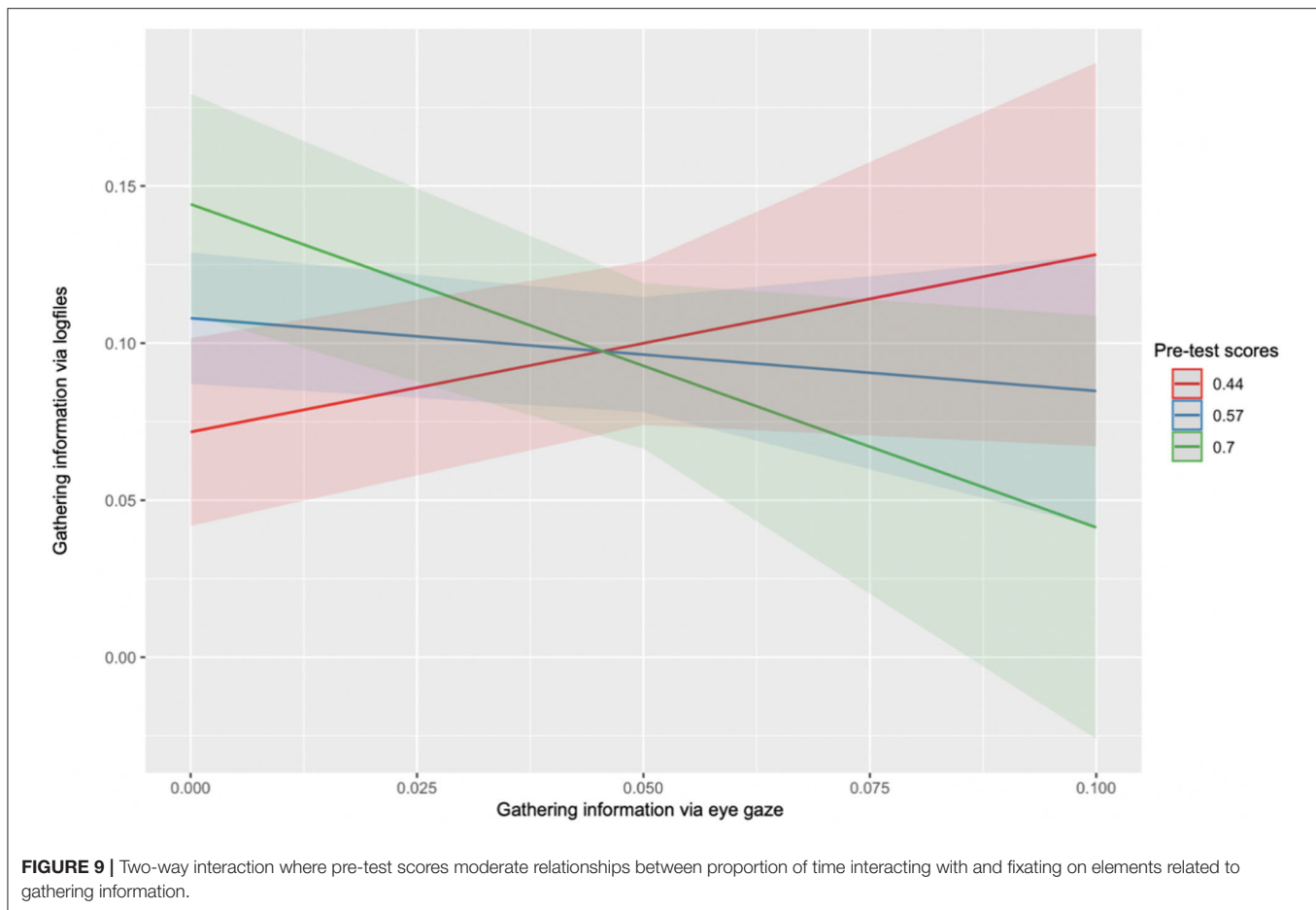
To examine whether relationships existed between interactions and fixations for elements related to non-scientific reasoning, while controlling for pre-test scores and experimental condition, we built a multiple linear regression equation. For this model, the fixations on elements was included as a predictor while interactions with elements was included as the target. Analyses suggested there were no significant relationships between interactions and fixations for elements related to non-scientific reasoning, while controlling for pre-test scores and experimental condition ($ps > 0.05$).

4. DISCUSSION

In this study, we examined scientific thinking with GBLEs to determine optimal features for individualizing instruction during science learning. Specifically, the objective of our study was to examine whether learners' multichannel data generated during game-based learning with Crystal Island were related to scientific thinking and performance and could be used to guide individualized instruction with GBLEs.

TABLE 6 | Modeling in-game actions related to scientific reasoning; * $p = 0.05$; *** $p < 0.05$.

Predictor variables	β	Standard error	t
Information gathering AOI	—	—	—
Condition	-0.03***	0.12	-0.22
Pre-test scores	—	—	—
Pre-test*Condition	—	—	—
Pre-test*Information AOI	-0.26***	0.12	-2.20
Adjusted R^2		0.189	
F		3.235***	
Hypothesis generation AOI	0.96	0.53	1.807
Condition	-0.02***	0.04	-0.425
Pre-test scores	—	—	—
Pre-test*Condition	—	—	—
Pre-test*Hypothesis AOI	—	—	—
Adjusted R^2		0.113	
F		2.23***	



4.1. Research Question 1

Our first research question examined relationships between fixations and interactions for game elements related to scientific reasoning and post-test scores, while controlling for pre-test scores. The predictive models suggested there were no relationships between fixation variables related to scientific reasoning and post-test scores ($p > 0.05$). However, the predictive models did reveal a significant three-way interaction between pre-test scores, interactions with game elements related to gathering information and generating hypotheses during game-based learning, and post-test scores. Specifically, the model suggested interactions with game elements related to gathering information and generating hypotheses moderated relationships between pre- and post-test scores. We found that when holding interactions for game elements related to gathering information and generating hypotheses constant at various increments—i.e., at mean $- 1$ SD, mean, mean $+ 1$ SD; see **Figure 8**), the relationships between pre- and post-test scores changed. For example, when learners spent more time, or an average proportion of time generating hypotheses, the proportion of time they spent gathering information positively influenced relationships between pre- and post-test scores. However, when the learner spent more time gathering information and less time generating hypotheses, the relationships between pre- and

post-test scores were no longer significant ($p > 0.05$). This finding suggested that when learners spend a higher proportion of time gathering information and less time generating hypotheses during game-based learning captured by combining both eye-gaze and in-game actions, it was detrimental to performance regardless of how much prior knowledge learners had brought to the learning session. These findings are partially consistent with our hypothesis such that there were relationships between interactions for game elements related to scientific reasoning and post-test scores based on empirical studies (Lazonder and Harmsen, 2016, p. 681; Alonso-Fernández et al., 2019, p. 301; Giannakos et al., 2019, p. 108; Sharma et al., 2020, p. 480) and SDDS theory (Klahr and Dunbar, 1988, p. 1); however, since no relationships between fixations on game elements related to scientific reasoning alone and post-test scores were found, this was inconsistent with previous literature (Alonso-Fernández et al., 2019, p. 301). We would like to emphasize that interaction data were *only* included in our models if eye-gaze data suggested that while learners were interacting with elements via log files, they were also fixating on those same elements (see section 2.6.2). While we did not find relationships when only including eye-gaze data, a possible explanation could be that when assessing eye-gaze data alone (i.e., without considering interaction data), it is too granular to capture its relation to performance. Instead,

other modalities of data generated during game-based learning may need to supplement statistical models in order to detect an effect or relationship, which has been a consistent finding across a number of GLA studies (Giannakos et al., 2019, p. 108; Sharma et al., 2020, p. 480). Future research should look toward including data channels such as concurrent verbalizations (Greene et al., 2018, p. 323). For instance, proportion of time individual learners' fixate and interact with game elements related to different components of scientific thinking that are supplemented by think-alouds could be useful in informing GLA that is individualized to each learner's needs.

4.2. Research Question 2

Our second research question examined whether the proportion of time fixating on game elements related to scientific reasoning predicted the proportion of time interacting with game elements related to scientific reasoning, while controlling for pre-test scores and experimental condition. The analyses revealed significant relationships between pre-test scores as well as interactions and fixations with elements related to gathering information. Specifically, we found a two-way interaction where the predictive model revealed that learners' prior knowledge about microbiology moderated relationships between interactions and fixations with game elements related to gathering information, such that the less prior knowledge learners had upon entering the learning session (e.g., <50% items correct on the pre-test assessment), then there was a positive relationship between interactions and fixations for elements related to gathering information. However, if the learner had scored higher on the pre-test (e.g., more than 50% items correct on the pre-test assessment), there was a negative relationship between interactions and fixations on games elements related to gathering information. We also found significant positive relationships between experimental conditions as well as interactions and fixations for game elements related to generating hypotheses. This finding suggests that learners in the partial agency condition spent significantly less time interacting with game elements related to generating hypotheses compared to the total agency condition. The predictive model also suggested that, while controlling for experimental condition, there was a significant positive relationship between interactions and fixations for game elements related to generating hypotheses during game-based learning. Unfortunately, we did not find significant relationships between interactions and fixations for elements related to experimental testing ($p > 0.05$). It is also important to note that we built separate models for each scientific reasoning variable (e.g., gathering information or generating hypotheses) due to multicollinearity issues. However, multicollinearity was not an issue for non-scientific reasoning variables, suggesting that combining eye-gaze and in-game actions data might provide insight into the extent to which a learner is engaging in scientific reasoning. In other words, if eye-gaze and in-game actions data are unrelated for certain game elements (e.g., non-scientific reasoning elements), it might suggest that the learner is not engaging in scientific reasoning.

These findings are partially consistent with our hypothesis. Specifically, significant relationships between pre-test scores as

well as interactions and fixations for game elements related to gathering information was consistent with our hypothesis, empirical literature (Hillaire et al., 2009, p. 43; Huang et al., 2015, p. 1049; Park et al., 2016, p. 796; Singh et al., 2018, p. 488), and SDDS framework (Klahr and Dunbar, 1988, p. 1), where we expected eye-gaze and prior knowledge to predict scientific reasoning in-game actions with GBLEs. Additionally, significant relationships between experimental condition as well as interactions and fixations for elements related to generating hypotheses was consistent with our hypothesis and empirical literature (Hillaire et al., 2009, p. 43; Huang et al., 2015, p. 1049; Park et al., 2016, p. 796; Singh et al., 2018, p. 488; Taub et al., 2020b, p. 1). However, the findings did not reveal significant relationships between interactions and fixations for elements related to experimental testing, which is inconsistent with our hypothesis and empirical evidence (Hillaire et al., 2009, p. 43; Huang et al., 2015, p. 1049; Park et al., 2016, p. 796; Singh et al., 2018, p. 488). A possible explanation of this result could be that, while fixations and interactions for elements were related for gathering information and generating hypotheses game elements, the game elements defined as experimental testing such as the scanner for testing food items did not require the learners to interact with as much as the game elements defined for gathering information and generating hypotheses. For instance, learners may have spent the majority of their time during game-based learning reading about the content and finding clues for formulating their hypotheses, and so upon testing hypotheses, it only required a few seconds or minutes to test those food items, reducing the proportion of time learners could have fixated and interacted with those game elements. As such, it is critical for researchers, educational technologists, and instructional designers to critically examine and build game elements that are not biased toward one component of scientific thinking relative to another in order to capture the nuances of scientific thinking. For example, Crystal Island has a disproportionate amount of game elements that learners can fixate on and interact with that reflect gathering information, such as the learner gathering clues from posters, research articles, books, and non-player characters. Or providing multiple tools that learners can use to track their clues and generate hypotheses during problem solving, yet there are few tools available for testing those hypotheses. Because of this, it is imperative to design GBLEs as both training and research tools and critically evaluate the design of GBLEs to capture dynamic learning behaviors that not only support GLA, but are also based on theoretical frameworks and empirical evidence.

4.3. Research Question 3

Our third research question examined whether relationships existed between interactions and fixations for game elements related to non-scientific reasoning, while controlling for pre-test scores. Our analyses revealed no significant relationships between interactions and fixations for game elements related to non-scientific reasoning. These findings were inconsistent with our hypotheses and previous empirical literature (Hillaire et al., 2009, p. 43; Huang et al., 2015, p. 1049; Park et al., 2016, p. 796; Singh et al., 2018, p. 488), where we expected

there to be relationships between interactions and fixations for game elements related to non-scientific reasoning. A possible explanation of this could be related to the objective of Crystal Island. Learners were required to solve a mysterious illness in order to complete the game and so they had no reason to fixate on or interact with game elements unrelated to scientific reasoning, or completing the objective of the game such as fixating and interacting with a plant or furniture during game-based learning. As such, a disproportionately low volume of data were present for fixations and interactions with elements related to non-scientific reasoning game elements, which may not have provided enough variance to detect a relationship in the predictive models. These findings highlight how critical it is to account for the objective of a learning environment when using GLA to guide instructional decision making and gain insight into complex learning behaviors.

4.4. Limitations

Limitations of this research include sampling bias such that the participants who completed the study were all undergraduate students and so the modeling results may not generalize to other learners who are not undergraduate students. Additionally, the sample was exclusively young adults, restricting our modeling interpretations to learners ranging from ages 18 to 25. Additionally, the shortcomings associated with using stepwise regression for model selection include biased parameter estimation, among others (Harrell, 2015, chapter 4). We also had significantly less volume of data for fixations and interactions with non-scientific reasoning related game elements. We also did not analyze information on learners' background such as academic major, race, etc. The findings from this study with Crystal Island may also not generalize to other game-based learning environments.

4.5. Implications and Future Directions

Future research should examine the effect of modifying the degree of agency during game play based on multichannel data generated from individual learners that depends on several issues such as prior knowledge, background such as major, race, age, etc., proportion of time fixating and interacting with game elements related to scientific thinking, efficaciousness in using scientific-thinking skills, developing competency etc. to guide individualized game-learning analytics. For instance, developing a temporal threshold that monitors how often and how long learners are fixating and interacting with game elements related to scientific reasoning to identify whether there is a need for intervention could be a novel approach for addressing questions related to how and what to adapt during game-based learning. If a learner demonstrates high prior knowledge about the content, then affording the learner more agency seems appropriate, but what if the learner is unable to apply their knowledge during game-based learning to think scientifically? Based on the learners' multichannel data gathered during game-based learning, GLA should be used in real-time to inform whether more or less agency is needed to scaffold scientific thinking and optimize individual learners' performance. Studies should also examine other data channels (e.g., concurrent verbalizations, physiology,

facial expressions of emotions) to assess whether these sources of information could supplement eye-gaze, in-game actions, and pre-test scores in their relation to scientific thinking and performance with GBLEs (Plass et al., 2020b). Implications of our findings lay a foundation for using multichannel data to define and capture scientific thinking during game-based learning that is crucial for supporting and optimizing individual scientific thinking, learning, and performance. Specifically, our study provides suggestions for building GBLEs that leverage agency as scaffolding techniques intended to, not only foster scientific-thinking skills, but capture, adapt, and inform instructional decision making based on the needs of individuals' scientific thinking during game-based learning. Further, to implement GLA in the classroom to inform instructional decision making, it will require more support and technological resources than are currently available to most educators, such as a teacher dashboard, to illustrate the various data channels and likely aid in making sense of the GLA and determining the appropriate intervention for individual learners.

5. CONCLUSIONS

As our world and communities are faced with persisting, and sometimes novel, socioeconomic, environmental, and health-related issues (e.g., prolonged interruptions in classroom instruction due to the COVID-19 pandemic), it is essential to equip learners of future generations with an interdisciplinary set of skills which contribute to higher-order thinking (e.g., scientific thinking; National Research Council, 2012; NASEM, 2018). The objective of our study was to examine relationships between scientific thinking and performance during game-based learning using learners' multichannel data—i.e., eye gaze, in-game actions, level of agency, and pre-test scores and post-test scores. Results showed significant predictive relationships between eye-gaze, pre-test scores, and interaction data related to scientific reasoning, suggesting that eye-gaze, prior knowledge, and agency play a crucial role in scientific thinking and performance with GBLEs. Overall, our findings highlight how critical it is to capture multichannel data, specifically a combination of in-game actions, eye gaze, degree of agency, and pre-test scores, when investigating scientific thinking with GBLEs. Using multiple data channels to define scientific thinking has the potential to support individualized instructional decision making that is guided by a combination of the learners' level of prior knowledge about the topic, as well as how long and what learners are fixating on and interacting with in terms of their relation to scientific reasoning. For instance, since the models revealed that prior knowledge moderates relationships between fixations and interactions with game elements related to gathering information, then a learner's pre-test score should inform the researcher, instructor, or system about if and when they should intervene based on whether the individual learner is demonstrating GLA which may be detrimental to learning and performance. Perhaps, the learner has low prior knowledge about microbiology upon entering the learning session, and so this would inform the instructional decision being made. If the

learner is not fixating on game elements related to gathering information, which is critical for their scientific thinking (i.e., hypothesis generation) when there is little prior knowledge about microbiology, then the researcher, instructor, or system would be aware that these data suggest the learner may not be engaging in scientific-thinking actions. Yet, if the learner has a relatively high prior knowledge about the content, then the researcher, instructor, or system need not worry about whether the learner is fixating on game elements related to gathering information as it is not critical for their scientific-thinking actions. To implement GLA in the classroom to assist in instructional decision making, educators will need more support and technological resources to help effectively monitor and understand learners' data. Recent studies have begun investigating dashboards as tools in the classroom to assist in data sense-making for guiding instruction across individual learners (Perez-Colado et al., 2017, p. 51). Additionally, our findings draw us to evaluate the role of agency designed into Crystal Island when referring to the three-way interaction. In other words, was one experimental condition designed to require learners to engage in one component of scientific reasoning more than another and why (i.e., why require learners in partial agency to interact with game elements related to gathering information compared to generating hypotheses)? Since our preliminary analyses suggested that learners assigned to the partial agency condition spent a significantly higher proportion of time interacting with game elements related to gathering information compared to the total agency condition, while the total agency condition spent a significantly higher proportion of time interacting with game elements related to generating hypotheses compared to the partial agency condition, it suggests that the degree of agency afforded to learners serves as a critical scaffolding technique that impacts scientific thinking and performance (Taub et al., 2020b, p. 1). As such, it is our responsibility as researchers and scientific thinkers to critically evaluate what each condition is requiring learners to do and why? Perhaps, designing levels of agency in GBLEs should be based on the prior knowledge the learner brings to the session to optimize their scientific thinking, learning, and performance. In the progressing realm of interconnected human and technology

ecosystems, implications of our findings could enhance our understanding of scientific thinking during game-based learning to effectively address future GBLEs that are designed for individualized and adaptive instruction to optimize scientific thinking, learning, and performance.

DATA AVAILABILITY STATEMENT

The datasets generated for this study will be made available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by IRB at North Carolina State University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RA designed and conducted the study. EC, DD, and MW wrote the first draft of the manuscript. DD, MW, and RA provided several rounds of edits on the manuscript. All authors contributed to the conception of the work and revised the final manuscript.

FUNDING

This study was supported by funding from the Social Sciences and Humanities Research Council of Canada (SSHRC 895-2011-1006). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of SSHRC.

ACKNOWLEDGMENTS

The authors would like to thank team members of the SMART Lab and Hans Doderlein at the University of Central Florida as well as the Intellimedia group at NC State for their assistance and contributions.

REFERENCES

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Control* 19, 716–723. doi: 10.1109/TAC.1974.1100705
- Alonso-Fernández, C., Cano, A. R., Calvo-Morata, A., Freire, M., Martínez-Ortiz, I., and Fernández-Manjón, B. (2019). Lessons learned applying learning analytics to assess serious games. *Comput. Hum. Behav.* 99, 301–309. doi: 10.1016/j.chb.2019.05.036
- Bandura, A. (2001). Social cognitive theory: an agentic perspective. *Annu. Rev. Psychol.* 52, 1–26. doi: 10.1146/annurev.psych.52.1.1
- Bondareva, D., Conati, C., Feyzi-Behnagh, R., Harley, J. M., Azevedo, R., and Bouchet, F. (2013). “Inferring learning from gaze data during interaction with an environment to support self-regulated learning,” in *International Conference on Artificial Intelligence in Education*, eds H. C. Lane, K. Yacef, J. Mostow, and P. Pavlik (New York, NY: Springer), 229–238. doi: 10.1007/978-3-642-39112-5_24
- Byrnes, J. P. and Dunbar, K. N. (2014). The nature and development of critical-analytic thinking. *Educ. Psychol. Rev.* 26, 477–493. doi: 10.1007/s10648-014-9284-0
- Center, S. E. (2014). 3. 4. 165 [apparatus and software]. *SensoMotoric Instruments*. Boston, MA.
- Clark, D. B., Tanner-Smith, E. E., and Killingsworth, S. S. (2016). Digital games, design, and learning: a systematic review and meta-analysis. *Rev. Educ. Res.* 86, 79–122. doi: 10.3102/0034654315582065
- Deater-Deckard, K., Chang, M., and Evans, M. E. (2013). Engagement states and learning from educational games. *New Direct. Child Adolesc. Dev.* 2013, 21–30. doi: 10.1002/cad.20028
- Dunbar, K. N., and Klahr, D. (2012). *Scientific Thinking and Reasoning*. Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780199734689.013.0035
- Freire, M., Serrano-Laguna, Á., Manero, B., Martínez-Ortiz, I., Moreno-Ger, P., and Fernández-Manjón, B. (2016). “Game learning analytics: learning analytics for serious games,” in *Learning, Design, and Technology*, eds M. J. Spector, B. B. Lockee, and M. D. Childress (Cham: Springer Nature), 1–29. doi: 10.1007/978-3-319-17727-4_21-1
- Giannakos, M. N., Sharma, K., Pappas, I. O., Kostakos, V., and Velloso, E. (2019). Multimodal data as a means to understand the learning experience. *Int. J. Inform. Manage.* 48, 108–119. doi: 10.1016/j.ijinfomgt.2019.02.003

- Greene, J. A., Deekens, V. M., Copeland, D. Z., and Yu, S. (2018). *Capturing and Modeling Self-Regulated Learning Using Think-Aloud Protocols, 2nd Edn.* New York, NY: Routledge; Taylor & Francis. doi: 10.4324/9781315697048-21
- Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics* 11, 1–21. doi: 10.1080/00401706.1969.10490657
- Harrell, F. E. (2015). “Multivariable modeling strategies,” in *Regression Modeling Strategies* (Cham: Springer), 63–102. doi: 10.1007/978-3-319-19425-7_4
- Hillaire, S., Lécuyer, A., Breton, G., and Corte, T. R. (2009). “Gaze behavior and visual attention model when turning in virtual environments,” in *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology* (New York, NY: ACM), 43–50. doi: 10.1145/1643928.1643941
- Huang, C.-M., Andrist, S., Sauppé, A., and Mutlu, B. (2015). Using gaze patterns to predict task intent in collaboration. *Front. Psychol.* 6:1049. doi: 10.3389/fpsyg.2015.01049
- iMotions (2014). *Attention Tool Version 5.4 [Computer Software]*.
- Johnson, P. O., and Neyman, J. (1936). Tests of certain linear hypotheses and their application to some educational problems. *Stat. Res. Memoirs* 1, 57–93.
- Klahr, D., and Dunbar, K. (1988). Dual space search during scientific reasoning. *Cogn. Sci.* 12, 1–48. doi: 10.1207/s15516709cog1201_1
- Klahr, D., Zimmerman, C., and Matlen, B. J. (2019). *Improving Students’ Scientific Thinking*. Cambridge University Press. doi: 10.1017/9781108235631.005
- Kuhl, P. K., Liang, S.-S., Guerriero, S., and Van Damme, D. (2019). *Developing Minds in the Digital Age: Towards a Science of Learning for 21st Century Education*. ERIC. doi: 10.1787/562a8659-en
- Lang, C., Siemens, G., Wise, A., and Gasevic, D. (2017). *Handbook of Learning Analytics*. Paris: SOLAR, Society for Learning Analytics and Research. doi: 10.18608/hla17
- Lazonder, A. W., Hagemans, M. G., and De Jong, T. (2010). Offering and discovering domain information in simulation-based inquiry learning. *Learn. Instruct.* 20, 511–520. doi: 10.1016/j.learninstruc.2009.08.001
- Lazonder, A. W., and Harmsen, R. (2016). Meta-analysis of inquiry-based learning: effects of guidance. *Rev. Educ. Res.* 86, 681–718. doi: 10.3102/0034654315627366
- Loderer, K., Pekrun, R., and Lester, J. C. (2020). Beyond cold technology: a systematic review and meta-analysis on emotions in technology-based learning environments. *Learn. Instruct.* 47, 1–15. doi: 10.1016/j.learninstruc.2018.08.002
- Long, J. A. (2019). *Interactions: Comprehensive, User-Friendly Toolkit for Probing Interactions*.
- Lüdecke, D. (2018a). sjmisc: Data and variable transformation functions. *J. Open Source Softw.* 3:754. doi: 10.21105/joss.00754
- Lüdecke, D. (2018b). sjplot: Data Visualization for Statistics in Social Science. New York, NY: R package version.
- Mayer, R. E. (2014). *Computer Games for Learning: An Evidence-Based Approach*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9427.001.0001
- McQuiggan, S. W., Goth, J., Ha, E., Rowe, J. P., and Lester, J. C. (2008). “Student note-taking in narrative-centered learning environments: individual differences and learning effects,” in *International Conference on Intelligent Tutoring Systems*, eds B. P. Woolf, E. Aïmeur, R. Nkambou, and S. Lajoie (Vienna: Springer), 510–519. doi: 10.1007/978-3-540-69132-7_54
- Miller Singley, A. T., and Bunge, S. A. (2018). Eye gaze patterns reveal how we reason about fractions. *Think. Reason.* 24, 445–468. doi: 10.1080/13546783.2017.1417909
- Morris, B., Croker, S., Zimmerman, C., Gill, D., and Romig, C. (2013). Gaming science: the “gamification” of scientific thinking. *Front. Psychol.* 4:607. doi: 10.3389/fpsyg.2013.00607
- Mulder, Y. G., Lazonder, A. W., and de Jong, T. (2011). Comparing two types of model progression in an inquiry learning environment with modelling facilities. *Learn. Instruct.* 21, 614–624. doi: 10.1016/j.learninstruc.2011.01.003
- Munoz, J., Yannakakis, G. N., Mulvey, F., Hansen, D. W., Gutierrez, G., and Sanchis, A. (2011). “Towards gaze-controlled platform games,” in *2011 IEEE Conference on Computational Intelligence and Games (CIG’11)*, (Berlin; Heidelberg: IEEE), 47–54. doi: 10.1109/CIG.2011.6031988
- NASEM (2018). *How People Learn II: Learners, Contexts, and Cultures*. Washington, DC: National Academies Press.
- National Research Council (2011). *Learning Science Through Computer Games and Simulations*. National Academies Press.
- National Research Council (2012). *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*. Washington, DC: National Academies Press.
- Park, H., Lee, S., Lee, M., Chang, M.-S., and Kwak, H.-W. (2016). Using eye movement data to infer human behavioral intentions. *Comput. Hum. Behav.* 63, 796–804. doi: 10.1016/j.chb.2016.06.016
- Perez-Colado, I. J., Perez-Colado, V. M., Freire-Moran, M., Martinez-Ortiz, I., and Fernandez-Manjon, B. (2017). “Integrating learning analytics into a game authoring tool,” in *International Conference on Web-Based Learning*, eds H. Xie, E. Popescu, G. Hancke, and M. B. Fernández (New York, NY; Cham: Springer), 51–61. doi: 10.1007/978-3-319-66733-1_6
- Peterson, R. A., and Cavanaugh, J. E. (2019). Ordered quantile normalization: a semiparametric transformation built for the cross-validation era. *J. Appl. Stat.* 47, 1–16. doi: 10.1080/02664763.2019.1630372
- Plass, J. L., Homer, B. D., and Kinzer, C. K. (2015). Foundations of game-based learning. *Educ. Psychol.* 50, 258–283. doi: 10.1080/00461520.2015.1122533
- Plass, J. L., Homer, B. D., Mayer, R. E., and Kinzer, C. K. (eds.). (2020a). *Theoretical Foundations of Game-Based and Playful Learning*. Cambridge, MA: MIT Press.
- Plass, J. L., Mayer, R. E., and Homer, B. D. (eds.). (2020b). *Handbook of Game-Based Learning*. Cambridge, MA: MIT Press.
- Plummer, P., DeWolf, M., Bassok, M., Gordon, P. C., and Holyoak, K. J. (2017). Reasoning strategies with rational numbers revealed by eye tracking. *Attent. Percept. Psychophys.* 79, 1426–1437. doi: 10.3758/s13414-017-1312-y
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna.
- Rajendran, R., Kumar, A., Carter, K. E., Levin, D. T., and Biswas, G. (2018). “Predicting learning by analyzing eye-gaze data of reading behavior,” in *International Conference on Educational Data Mining* eds K. E. Boyer and M. Yudelson (New York, NY: ERIC), 455–461.
- Raven, J. (2000). The raven’s progressive matrices: change and stability over culture and time. *Cogn. Psychol.* 41, 1–48. doi: 10.1006/cogp.1999.0735
- Raven, J. C. (1941). Standardization of progressive matrices. *Br. J. Med. Psychol.* 19, 137–150. doi: 10.1111/j.2044-8341.1941.tb00316.x
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Q. J. Exp. Psychol.* 62, 1457–1506. doi: 10.1080/17470210902816461
- Rowe, J. P., Shores, L. R., Mott, B. W., and Lester, J. C. (2011). Integrating learning, problem solving, and engagement in narrative-centered learning environments. *Int. J. Artif. Intell. Educ.* 21, 115–133. doi: 10.3233/JAI-2011-019
- Salvucci, D. D., and Goldberg, J. H. (2000). “Identifying fixations and saccades in eye-tracking protocols,” in *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* (New York, NY: Association for Computing Machinery), 71–78. doi: 10.1145/355017.355028
- Scheiter, K., and Eitel, A. (2017). “The use of eye tracking as a research and instructional tool in multimedia learning,” in *Eye-tracking Technology Applications in Educational Research*, eds C. Was, F. Sansosti, and B. Morris (Hershey, PA: IGI Global), 143–164. doi: 10.4018/978-1-5225-1005-5.ch008
- Sharma, K., Papamitsiou, Z., Olsen, J. K., and Giannakos, M. (2020). “Predicting learners’ effortful behaviour in adaptive assessment using multimodal data,” in *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (New York, NY: Association for Computing Machinery), 480–489. doi: 10.1145/3375462.3375498
- Shute, V., Rahimi, S., and Lu, X. (2019). “Supporting learning in educational games: promises and challenges,” in *Learning in a Digital World*, eds P. Diaz, A. Ioannou, K. Bhagat, and J. Spector (Springer), 59–81. doi: 10.1007/978-981-13-8265-9_4
- Shute, V. J., Wang, L., Greiff, S., Zhao, W., and Moore, G. (2016). Measuring problem solving skills via stealth assessment in an engaging video game. *Comput. Hum. Behav.* 63, 106–117. doi: 10.1016/j.chb.2016.05.047
- Singh, R., Miller, T., Newn, J., Sonenberg, L., Velloso, E., and Vetere, F. (2018). “Combining planning with gaze for online human intention recognition,” in *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems* (Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems), 488–496.
- Smith, G., Shute, V., and Muenzenberger, A. (2019). Designing and validating a stealth assessment for calculus competencies. *J. Appl. Test. Technol.* 20, 52–59.
- Taub, M., Azevedo, R., Bradbury, A. E., Millar, G. C., and Lester, J. (2018). Using sequence mining to reveal the efficiency in scientific reasoning during stem

- learning with a game-based learning environment. *Learn. Instruct.* 54, 93–103. doi: 10.1016/j.learninstruc.2017.08.005
- Taub, M., Mudrick, N., Bradbury, A. E., and Azevedo, R., editors (2020a). *Self-Regulation, self-Explanation, and Reflection in Game-Based Learning*. Cambridge, MA: MIT Press.
- Taub, M., Sawyer, R., Smith, A., Rowe, J., Azevedo, R., and Lester, J. (2020b). The agency effect: the impact of student agency on learning, emotions, and problem-solving behaviors in a game-based learning environment. *Comput. Educ.* 147, 1–19. doi: 10.1016/j.compedu.2019.103781
- van Zoest, W., Van der Stigchel, S., and Donk, M. (2017). Conditional control in visual selection. *Attent. Percept. Psychophys.* 79, 1555–1572. doi: 10.3758/s13414-017-1352-3
- Venables, W., and Ripley, B. (2002). “Random and mixed effects,” in *Modern Applied Statistics With S* (New York, NY: Springer), 271–300. doi: 10.1007/978-0-387-21706-2_10
- Vendetti, M. S., Starr, A., Johnson, E. L., Modavi, K., and Bunge, S. A. (2017). Eye movements reveal optimal strategies for analogical reasoning. *Front. Psychol.* 8:932. doi: 10.3389/fpsyg.2017.00932
- Wickham, H., et al. (2007). Reshaping data with the reshape package. *J. Stat. Softw.* 21, 1–20. doi: 10.18637/jss.v021.i12
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis, 2nd Edn.* New York, NY: Springer. doi: 10.1007/978-3-319-24277-4_9
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., et al. (2019). Welcome to the tidyverse. *J. Open Source Softw.* 4:1686. doi: 10.21105/joss.01686
- Wickham, H. and Bryan, J. (2017). *readxl: Read excel files. R package version 1.0.0.* Available online at: <https://CRAN.R-project.org/package=readxl>
- Wickham, H., François, R., Henry, L., and Müller, K. (2018). *dplyr: A Grammar of Data Manipulation.* R Package Version 0.7.6.
- Wouters, P., Van Nimwegen, C., Van Oostendorp, H., and Van Der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *J. Educ. Psychol.* 105, 249–265. doi: 10.1037/a0031311
- Zimmerman, C., and Croker, S. (2014). A prospective cognition analysis of scientific thinking and the implications for teaching and learning science. *J. Cogn. Educ. Psychol.* 13, 245–257. doi: 10.1891/1945-8959.13.2.245

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Cloude, Dever, Wiedbusch and Azevedo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.