International Journal of Serious Games

ISSN: 2384-8766
https://journal.seriousgamessociety.org/

Article

From product to process data: Game mechanics for science learning[1]

Daryn Dever[1], Megan Wiedbusch[2], Cameron Marano[2], Annamarie Brosnihan[2], Kevin Smith[2], Milouni Patel[2], Tara Delgado[2], James Lester[3], and Roger Azevedo[2]

[1]College of Education, University of Florida, Gainesville, USA; [2]School of Modeling, Simulation, and Training, University of Central Florida, Orlando, USA; [3]College of Engineering, North Carolina State University, Orlando, USA
{ddever} @ufl.edu ; {megan.wiedbusch, Cameron.marano, Annamarie.brosnihan, kevin.smith4, milouni.patel, tara.delgado, roger.azevedo} @ucf.edu; {lester}@ncsu.edu

Keywords:

Game-based Learning
Transition Matrices
Hierarchical Clustering

Received: May 2024
Accepted: October 2024
Published: November 2024
DOI: 10.17083/ijsg.v11i4.790

Abstract

Game-based learning environments (GBLEs) supplement classroom instruction so students can demonstrate their scientific reasoning abilities and increase knowledge, providing a platform that promotes interest and engagement in science. The goal of this study was to examine the effectiveness of game mechanics for science learning. This study identifies how two types of game mechanics—learning and assessment mechanics—are used by high school participants (N = 137) as they learn about microbiology with *Crystal Island*, a game-based learning environment for science education. Participants' learning outcomes were evaluated in two ways: learning gains, which assessed participants' domain knowledge acquisition, and game completion, which assessed participants' ability to successfully demonstrate scientific reasoning abilities. Results from this study showed that game completion is not related to learning gains. However, as participants engaged with increasingly more assessment mechanics, learning gains decreased. Further, profiles of learners were extracted to better understand the learning process that best supports greater learning outcomes. Results showed that learners who engaged in less recurrent transitions across assessment mechanics were more likely to successfully demonstrate scientific reasoning abilities. Implications for the design of games which provide scaffolding based on process data of learners' game mechanic use are provided.

[1] An earlier version of this paper was presented at the GALA 2023 conference (Dublin, Ireland) and published as Dever, D. A., Wiedbusch, M., Park, S., Llinas, A., Lester, J., & Azevedo, R. (2024). Assessing the complexity of game mechanic use during science learning. *In Proceedings of the 12th Annual Games and Learning Alliance Conference* (pp. 299-308). Dublin, Ireland. Cham: Springer Nature Switzerland.

1. Introduction

Students who are about to enter higher education and the workforce are woefully unprepared in terms of achievement in science. A U.S. National Assessment of Educational Progress report card released in 2019 [1] showed that while nearly half of all twelfth-graders report that they are somewhat likely or more likely to pursue careers in science, 34% of the same population reported low interest and enjoyment of science. Further, 50% of twelfth-graders reported that they never or only occasionally participate in inquiry-related science activities in the classroom. This contributes to the 41% of students who are below a basic science proficiency level. The COVID-19 pandemic illuminated this lack of understanding in the health sciences where studies have shown that high-school students possess a significant lack of knowledge about COVID-19. This was related to their noncompliance with public health policies and engagement in risky behaviors that threaten disease spread [2-3].

The goal of this study is to explore the use of games for science learning by studying learning not just as a product, such as that assessed via standardized testing, but as a set of processes that temporally unfold over time and are dependent upon the individual learner characteristics, context and affordances of the game environment. To achieve this goal, the current study examines how high-school learners sequentially interact with different game mechanics within a game-based learning environment and how these interactions relate to increased learning gains and learners' ability to successfully complete game objectives.

1.1 Games for Science Learning

Literature has found game-based learning environments (GBLEs) to be effective tools for learning [4]–[8]. Relevant to our study are GBLEs focused on scientific literacy including knowledge about the domain as well as the use of scientific inquiry and reasoning. Several games, such as *Crystal Island* [9]-[12], *Operation ARA* [13], and *River City* [14]-[15], have situated learners as scientists to identify problems, collect information, and hypothesize to achieve some goal. Across all games, there are several studies which pose games for science learning as effective tools which aid all learners, including low-achieving students, in learning and demonstrating the scientific reasoning process [9]-[15]. However, some work has questioned the usefulness and effectiveness of these tools within an educational context, even questioning claims that ground the existence of games for learning [16]-[18]. One argument is that games allow for a continuous assessment of skills, not necessarily *teaching* those skills and knowledge, providing learners the ability to demonstrate skills within a somewhat realistic environment [19]. As such, it becomes essential to identify which affordances (i.e., game mechanics) within the GBLE are used to *teach* skills and which ones are used to *assess* skills so that learners' interactions with game mechanics can be tracked and related to learning outcomes with the ultimate goal of verifying the effectiveness of games for science learning. This paper responds to Gris and Bengtson's [16] call for a closer examination of the effectiveness of GBLEs by: (1) tracing how learners interact with different game mechanics throughout the learning process; and (2) evaluating learning outcomes on two spectrums—domain knowledge gained using pre- and post-tests demonstrating learned concepts as well as game completion success demonstrating scientific reasoning skills.

1.2 Theoretical Framework

GBLEs leverage the interest and motivation inherent within games to promote learning without sacrificing learners' affective, behavioral, cognitive, and social engagement [6]. As such, there is a balance that needs to be struck within GBLEs that equalizes the amount of interest that games for play offer and the progression of learning outcomes via instructional designs and materials. Plass et al.'s [6] framework for playful learning describes how design elements (i.e., game mechanics, visual design, sound design, and narrative design) interact with engagement types (i.e., behavioral, cognitive, affective, and sociocultural) to promote the learners' cognitive, affective, metacognitive, motivational, and sociocultural processes, thus increasing learning outcomes. Design elements refer to the affordances and features of a GBLE in which learners interact with elements directly (i.e., game mechanics, narrative design, or content and skills) or observe them throughout the game (i.e., visual aesthetic design and sound design; [6]). These elements are incorporated as a means of promoting an engaging, but educational, GBLE. This study specifically explores the role of game mechanics, categorized as either a learning or assessment game mechanics, in supporting learning outcomes. As the design elements, including game mechanics, are consequential to the targeted engagement and learning of a learner, the interactions (or lack thereof) between the learner and design elements can promote one's cognitive, affective, metacognitive, and motivational processes essential to learning [6]. While previous literature has focused on how design elements influence learning processes and engagement [20]-[23], the conception of how game mechanics interact with each other to promote a learning goal is unclear.

1.3 Game Mechanics

Game mechanics are one of the key pillars that support effective and engaging games for learning [6]. These mechanics are defined as the activities within a game repeatedly engaged with by the learner that represent the intersection of pedagogical practices used to support learning and the design aspects that promote play, interest, and engagement [6],[24]-[25]. It is important to note that game mechanics result in a series of behaviors enacted by the learner throughout the game, allowing researchers to identify when learners engage in these behaviors and which mechanics are used. Game mechanics can be divided into two types: learning mechanics, which have a learning focus, and assessment mechanics, which focus on assessing learners' domain understanding and goals. In this section, we review the two essential types of game mechanics for both learning domain knowledge and demonstrating knowledge and skills within the GBLE. It is important to note that the definitions and activities which fall under learning and assessment mechanics are based on the definitions from Plass et al.'s [6] integrative framework for playful learning. We acknowledge that other frameworks for game-based learning in educational contexts, notably that of Arnab and colleagues' [24] Learning Mechanics-Game Mechanics model for pedagogy and design, may define and categorize these terms differently. While this framework would have been a feasible option to situate our study in, the goal of our study was to acknowledge the cognitive processes which occurred for learners to gain domain knowledge and engage in scientific reasoning required for successful game completion. As Plass et al.'s [6] framework addresses the role of game mechanics in learners' deployment of cognitive processes, this framework was chosen to situate our study. Thus, we adopted the definitions of game mechanics within this paper. Definitions of each mechanic, along with operationalizations contextualized to this study (see Coding & Scoring), are provided below according to Plass et al.'s [6] framework.

1.3.1 Learning Mechanics

Game designers have long advocated for the development of novel mechanics aimed at engaging players while promoting learning outcomes [26], including solving puzzles that require critical thinking and undertaking quests that demand the application of scientific reasoning to progress into the gameplay. Within GBLEs, learning mechanics are foundational components that bridge educational content with gameplay, facilitating an engaging and effective learning experience [27]. Learning mechanics are the diverse and multifaceted rules, systems, and features within games that govern player interactions and progression, tailored to achieve specific learning outcomes. These mechanics transform the features and elements within the game into significant and impactful learning experiences. Grounded in pedagogical theories, including situated learning, cognitive apprenticeship theory, and self-regulated learning, learning mechanics are designed to promote knowledge acquisition, skill development, and cognitive engagement [24],[27].

Learning mechanics can be categorized into static, dynamic, and aid mechanics, each serving distinct pedagogical functions. *Static learning mechanics,* encompassing traditional educational resources like books, research articles, and posters, are integrated into GBLEs to provide foundational knowledge and theoretical context or facts during the gameplay. They are considered static as learner does not interact with the content outside of basic material absorption. That is, these are resources provided to the learner that could be found outside of a GBLE as basic content. *Dynamic learning mechanics* are those interactions with non-player characters (NPCs) and other game elements, facilitating experiential learning and application of theoretical knowledge in simulated environments. These elements are dynamic in that they require the learner to interact with them to uncover their instructional and domain content. These often mimic interactions or processes that could not be found in a static representation. *Aid learning mechanics*, including worksheets and note-taking tools, support metacognitive activities and reflection, essential for consolidating learning [28]. These are the equivalent to strategy tools that allow for a learner to externalize (meta)cognitive processes that are used during the learning process.

Learning mechanics are often the subject of much scrutiny when studying the impact of GBLEs. Many games' learning mechanics can be undermined by the "chocolate-covered broccoli" phenomenon [29]-[30], wherein poorly integrated educational content and supports may compromise the game's enjoyment, disrupt the flow, and reduce it to mere instructional software [31]. These types of GBLEs have often been described as tests or worksheets dressed up as games which drive disengagement and disinterest [32]. However, when learning mechanics are embedded within the game's narrative or amongst other game mechanics, these games may be more efficient learning technologies. For instance, Dever et al. [33] observed significant learning gains when learners conversed with embedded non-player characters, highlighting the efficacy of dynamic and aid mechanics in promoting learners' metacognition and domain learning.

Other studies have observed varying effects of learning mechanics that are more static in nature as they do not require input from the learner (i.e., the learner is a passive receiver of information) on learning outcomes. Studies have shown that the content held within static learning mechanics (i.e., instructional text versus diagrams) can have both positive and negative effects on learning outcomes [33]. In a recent systematic review for serious games [34], similar findings have been corroborated, which emphasized the enhancement of learning outcomes through game mechanics designed to support cognitive and metacognitive processes.

The GBLE literature also indicates another disparity in the effectiveness of learning mechanics, emphasizing the necessity for games that are not only informed by different learning theories but also adapted to the nuances of individual learners' needs [35]. The investigation into how learners interact and use learning mechanics is critical for understanding how effective these mechanics are in aiding learning outcomes in terms of greater domain knowledge attained via learning mechanics. By investigating how learners interact with these mechanics throughout the learning process, this study can hold critical implications for the design of GBLEs that are both educationally potent and engaging [36]. Further, learning mechanics are essential to GBLE research as they connect pedagogical theories with practical applications, enhancing engagement, and facilitating skill development in GBLEs. By focusing on how learning mechanics operationalize learning theories and support cognitive, affective, metacognitive, and motivational skills, our study contributes to both theoretical and practical understandings of effective educational game design.

1.3.2 Assessment Mechanics

Within GBLEs, learning mechanics are integrated with assessment mechanics, which apply testing theories through embedded diagnostic tools. In relation to the current study, assessment mechanics can target learning outcomes by pinpointing areas of (mis)understanding through questionnaires of domain knowledge, examining learners' interaction with instructional materials, and evaluating learners' ability to use the knowledge they have gained to successfully complete the game [4],[37].

GBLEs can assess learning outcomes through both direct and indirect measures. A recent systematic literature review of learning assessments during game-based learning has shown that most studies measure learning outcomes solely through direct measures using pretest-post-test results [16]. A small number of studies incorporated indirect measures of learning outcomes, such as the GBLE recognizing an increase in the learner's understanding of the concept. This indicates that there is a current gap in the literature in which studies do not use both direct and indirect measures of learning. Addressing this gap can offer a better understanding of how assessment mechanics within games can be modified to foster an increase in learning outcomes.

There is a limited number of studies assessing learning outcomes with game mechanics that examine how and when learners transition between game mechanics (e.g., learning and assessment). Using log files collected throughout gameplay, the current study aims to show the significance of examining how learners' use and sequential deployment of learning and assessment mechanics can determine learning outcomes and facilitate learners' understanding of microbiology and use of scientific reasoning skills. This study also argues that assessment mechanics can serve a dual purpose, not only do they provide feedback on learners' growing understanding of the content in the game-based environment but they also facilitate an understanding of the narrative of the game.

1.4 Evaluating Learning Outcomes in Games

Traditional assessments (e.g., quizzes and exams) for learning have several limitations due to their administration and measurement constraints. They are often administered after learning has occurred [38], in an ecologically-removed approach such as multiple choice [39], and fail to assess everything that has been taught in a classroom [40]. GBLEs, however, allow for trace data (e.g., log files of computer-human interactions) to capture learning as it is occurring within the simulated environments [28],[41].

Trace data are unobtrusive measures that can act as artifacts of learning processes in real time without interruption [28],[42], and can be used to identify sequences of behaviors or game actions [43]. The sequences can then be used to assess various learning or psychological processes [9],[44], create learner profiles [45], or describe learner behaviors to cyclically feedback into a game for individualized adaptivity [43]. Much research has examined the mapping of log files capturing traces of player behavior to various cognitive and metacognitive processes (e.g., orientation, making judgments, applying learning strategies) within computer-based learning environments such as intelligent tutoring systems [46].

The introduction of trace data and conceptualization of learning both as a product and a process, however, come with the added complications of assessing performance outside of rote memorization of domain knowledge captured through traditional means. This is further exacerbated by the often-competing goals that accompany GBLEs. That is, especially within narrative-driven GBLEs, there can be competing goals for the user to balance. GBLEs have instructional-based goals that encourage players to gain domain knowledge about the subject. However, GBLEs can also embed narrative based goals that motivate and engage the learner towards "winning" the game or making progress within the game. There is limited research comparing when certain game elements or mechanics encourage prioritizing one set of goals over another [47]. Some GBLE designers have strategically used game mechanics to be used as assessment mechanics (i.e., stealth assessments; [48]). Stealth assessments collect learning patterns and learning outcomes instantaneously during gameplay without disrupting the learning process [49], such as with self-reports that are prompted throughout gameplay [50]. The inclusion of stealth assessments ensures the environment remains enjoyable, informative, and instructional without the learner necessarily being aware that certain actions have importance and are being observed for assessment purposes [51]-[52].

Stealth assessments can capture the "easy-to-measure" metrics such as declarative knowledge [51] but they are also able to capture processes and products related to (meta)cognitive competencies throughout the game [53] while remaining a part of the gameplay [55]. Often, this is accomplished by completing a one-to-one mapping of action to psychological process [41],[55]. However, this approach has primarily been used to describe aggregate measures of player actions, neglecting the temporal impact of the chronology in which actions were taken. Knowing the order and how learners transition between different types of actions may reveal more nuances about the context in which learning occurred [43]. Sequence pattern mining has been previously explored within GBLEs (e.g., [44],[56]-[57]), however these techniques also remain primarily descriptive in nature and fail to account for the type of actions as they relate to either game or learning mechanics. As such, our study closes this gap by applying a new approach for understanding how the transitions between types of game mechanics influence learners' game and learning outcomes.

2. Current Study

2.1 *Crystal Island*: A GBLE for Microbiology

Crystal Island is a GBLE intended to teach K-12 students about microbiology, including what infectious diseases are and how they spread, by requiring students to engage in scientific reasoning as they gather information about different diseases, hypothesize about a disease that is infecting researchers on the virtual island, and providing a final diagnosis based on the evidence that was gathered [10]. This GBLE consists of a mystery narrative that garners students' interest while simultaneously guiding them to engage in scientific reasoning. From this first-person view, students embark on a journey to save a research team on a camp from a mysterious illness infecting the researchers on the island [10]. Learners solve the mystery by engaging with several mechanics that serve different purposes, including informing the learner about various types of illnesses, demonstrating symptomology of the mysterious illness, and assessing learners' emerging understanding of microbiology. Throughout the environment, there are several resources to develop knowledge about microbiology and apply that knowledge to achieve learning and game goals.

At the beginning of *Crystal Island*, there is a one-and-a-half-minute video explaining the premise and creating a sense of cruciality of the situation on the island in the learner. After the video, learners arrive at the research camp to complete a tutorial teaching them about the different mechanics found in the game and how to interact with objects and resources in the environment. After the tutorial, learners are released to explore the island and start their process for diagnosing the infected researchers. On the island, learners can either explore the camp or enter different buildings such as the infirmary, cafeteria, an NPC's home, and a laboratory to interact with several elements. Specifically, an in-game diagnosis worksheet that will assist in formulating learners' findings, posters on the wall provide information about different illnesses, NPCs act as either patients to convey recurring symptoms or give domain information (e.g., size of a bacteria versus a virus), food items can be collected and tested for diseases using the scanner, books and research articles can provide domain information, and concept matrices can assess learners' knowledge. These affordances are always available to participant throughout their time on task with access to these materials available within each building, except for the scanner for food items which is found in the laboratory. After participants have narrowed down what could be the potential illness, they fill out the final diagnosis section on the worksheet with the correct illness (salmonellosis or influenza), transmission source (egg, bread, or milk), and treatment (rest or vaccination) and return to the infirmary to inform the camp nurse. If the final diagnosis is incorrect, the nurse will identify the error and recommend the student to keep working; if the participant submits a fully correct diagnosis, the mystery is solved, and the game is complete [10].

2.1.1 Prior Works on Crystal Island

Over the past several years, *Crystal Island* has been a vehicle to dissect game mechanics in GBLEs [9],[21],[58]-[62]. Game mechanics have been of particular interest, investigated by multiple studies using *Crystal Island*. For example, Dever et al. [33] examined the relationship between restricting learners' agency and learners' engagement with different types of information text presentations (e.g., static books, dynamic conversations with non-player characters). This study found that restricting agency leads to higher learning gains and the time spent on learning mechanics had varying effects on domain knowledge acquisition where the increase in learning gains depended on the type of text presentation from which domain information was provided to the learner. Emerson et al. [63] measured performance by the efficiency of learners' actions in combination with their solution attempts. Specifically, this study examined how often learners tested their hypotheses within *Crystal Island* compared to the number of attempts at providing the correct diagnosis. Similarly, Taub et al. [53] also reviewed game mechanics but had a particular interest on understanding the relationship between the quality in which learners used learning mechanics and the success of learners' assessment mechanic use. Specifically, this study examined how learners' use of texts on domain information were related to the frequency in which they attempted to successfully complete in-game assessment of the information held within the texts. This study found that learners who had greater performance on the assessments demonstrated a lower number of text use but a greater frequency (i.e., participants re-read the same texts). Although these past studies have examined learners' use of learning mechanics [33], assessment mechanics [63], and the relationship between these mechanics [53], few studies across GBLE literature have used process data to examine how learning and assessment mechanics are deployed in conjunction with each other and related to learning outcomes regarding both learning gains and game completion success.

2.2 Research Questions & Hypotheses

The goal of this study[2] is to contribute to recent GBLE literature in closing the gap between the current interest and achievement of K-12 students in STEM fields and the expectations of the STEM workforce. Specifically, this study addresses limitations in GBLE literature by contributing to the body of *Crystal Island* studies that utilize process data to assess how learners engage in both

[2] An earlier version of this paper was presented at the GALA 2023 conference (Dublin, Ireland) and published as Dever, D. A., Wiedbusch, M., Park, S., Llinas, A., Lester, J., & Azevedo, R. (2024). Assessing the complexity of game mechanic use during science learning. *In Proceedings of the 12th Annual Games and Learning Alliance Conference* (pp. 299-308). Dublin, Ireland. Cham: Springer Nature Switzerland.

learning and assessment game mechanics and how these relate to multiple learning outcomes including the increase of scientific knowledge and the success in demonstrating scientific reasoning skills for game completion. To accomplish these goals, several research questions were formed and investigated.

Research Question 1**:** Do learning gains differ between groups of learners who successfully completed *Crystal Island* versus learners who did not successfully complete the game? For this question, we hypothesize that learners who are successfully able to complete the game demonstrate their scientific reasoning skills and therefore would have greater science learning gains [19].

Research Question 2**:** Are learning outcomes related to the frequency in which learners deployed learning versus assessment game mechanics? For this research question, we hypothesize that learners who engage in a greater frequency of both learning and assessment mechanics would demonstrate greater learning outcomes in terms of greater science learning gains on pre- and post-test quizzes of microbiology knowledge. This is hypothesized as learning mechanics contribute to pedagogical outcomes, such as increased domain knowledge, and in-game assessment mechanics can allow learners to gauge what they have or have not yet learned thereby allowing them the opportunity to re-engage with instructional materials [6],[37],[64]-[65]. We also hypothesize that learners who were successful in completing the game objectives would have a greater frequency of assessment mechanics, rather than learning mechanics because assessment mechanics are able to inform students of their progress on game objectives [19].

Research Question 3**:** Are we able to identify distinct clusters of learners based on their transition frequency across game mechanics? From past studies [21], we hypothesize that we will be able to identify learners who are similar in their transition frequency characteristics but we do not assume a specific hypothesis from theory regarding these clusters of learners.

Research Question 4**:** Do clusters of learners who differ in their frequency of game mechanic transitions vary in their learning gains and game success? We hypothesize that clusters of learners who vary in their frequency of game mechanic transitions will demonstrate differences in their learning outcomes where clusters may demonstrate more optimal transitions than other clusters based on prior literature regarding clustering techniques in *Crystal Island* [9] as well as on past studies on game mechanic use [33],[53],[63].

Research Question 5**:** Is there a difference in the frequency in which learning and assessment game mechanic subtypes are deployed by clustered learners? We hypothesize that clusters of learners varying in their frequency of game mechanic transitions will demonstrate differences in their game mechanic subtypes based on prior literature on *Crystal Island* [9],[33],[53],[63]. However, we do not propose a direction for the specific subtypes as not enough literature has been released regarding these game mechanic subtypes and the clusters have not yet been theoretically defined.

Research Question 6**:** How do clusters of learners transition across subtypes of learning and assessment game mechanics? Based on prior literature [21], we hypothesize that learners will demonstrate a wide range of transitions across subtypes of learning and assessment mechanics, but we do not provide a direction on the differences in those probabilities across game mechanics.

3. Methodology

3.1 Participants

Participants were recruited from a public North American high school. Participants' ages ranged from 14 to 18 years old (M_{age} = 15.5; SD_{age} = 1.05). High school students were recruited via emails and letters sent home to parents by school administrators. These emails and letters detailed the purpose and goals of the study along with consent forms to be signed by the student's parents. In total, 148 students were recruited (19.5% female), however, 11 participants were removed from analyses due to missing post-test scores. This resulted in a final dataset of 137 participants' data analyzed for this study. The study was run during school hours during regular class time, and due to this, participants were not paid for their involvement with the study.

3.2 Experimental Setup, Materials, Procedure, and GBLE

Participants completed the study during class time, permission provided by teachers and administrators. Upon entering the classroom, participants with signed consent forms were provided login information by researchers. The study was completed on school computers with all study materials accessible through an internet browser using the login information. Upon logging into the study site, the pre-task questionnaires were provided. This included questions about demographics and a 17-item pre-test on microbiology. After participants completed the pre-task questionnaires (approximately 15 minutes), participants were asked to wait until fellow participants were finished with the pre-task questionnaires. Researchers then provided a brief explanation of the study. Participants were told they had 60 minutes to play and complete *Crystal Island*. Gameplay time was restricted to ensure that all study materials were completed within the 90-minute class timeframe.

Participants were then asked to play *Crystal Island* (Figure 1). This included a brief tutorial of the actions and mechanics available to participants within the environment. During *Crystal Island*, students explored an island. Their goal was to identify what disease has infected researchers on the island. To do this they had to gather evidence and learn about the symptoms, biology, and how each possible disease works. Across the island, various non-player characters (NPCs) were available to talk to waiting to share their knowledge and context clues. Along with the NPCs, there are also books, posters, and articles among the different houses with information about microbiology concepts. Students also were provided with a diagnostic worksheet which allowed them to synthesize information about diseases and their symptoms. Students were also able to "scan" food items for diseases to help gather further evidence and identify diseases. To complete the game, students were required to diagnose what disease had infected the researchers, the cause of it, and how to treat it.

Log files were collected on the participant's behaviors during *Crystal Island* gameplay. These log files included timestamps of behaviors to detail the order in which participants interacted with different game mechanics, including events such as conversing with NPCs, reading books, posters, and research articles, editing the worksheet, completing concept matrices, scanning food items for diseases, and submitting the final diagnosis.

After the 60-minutes time limit had passed, participants were instructed to stop and continue to the post-task questionnaires regardless of whether they finished or not. Participants who finished the task prior to the 60-minute time limit were automatically transferred to post-task questionnaires. Post-task questionnaires included another 17-item test on microbiology which was similar but not identical to the pre-test and an evaluation of workload and interest. Participants were then thanked for their time.

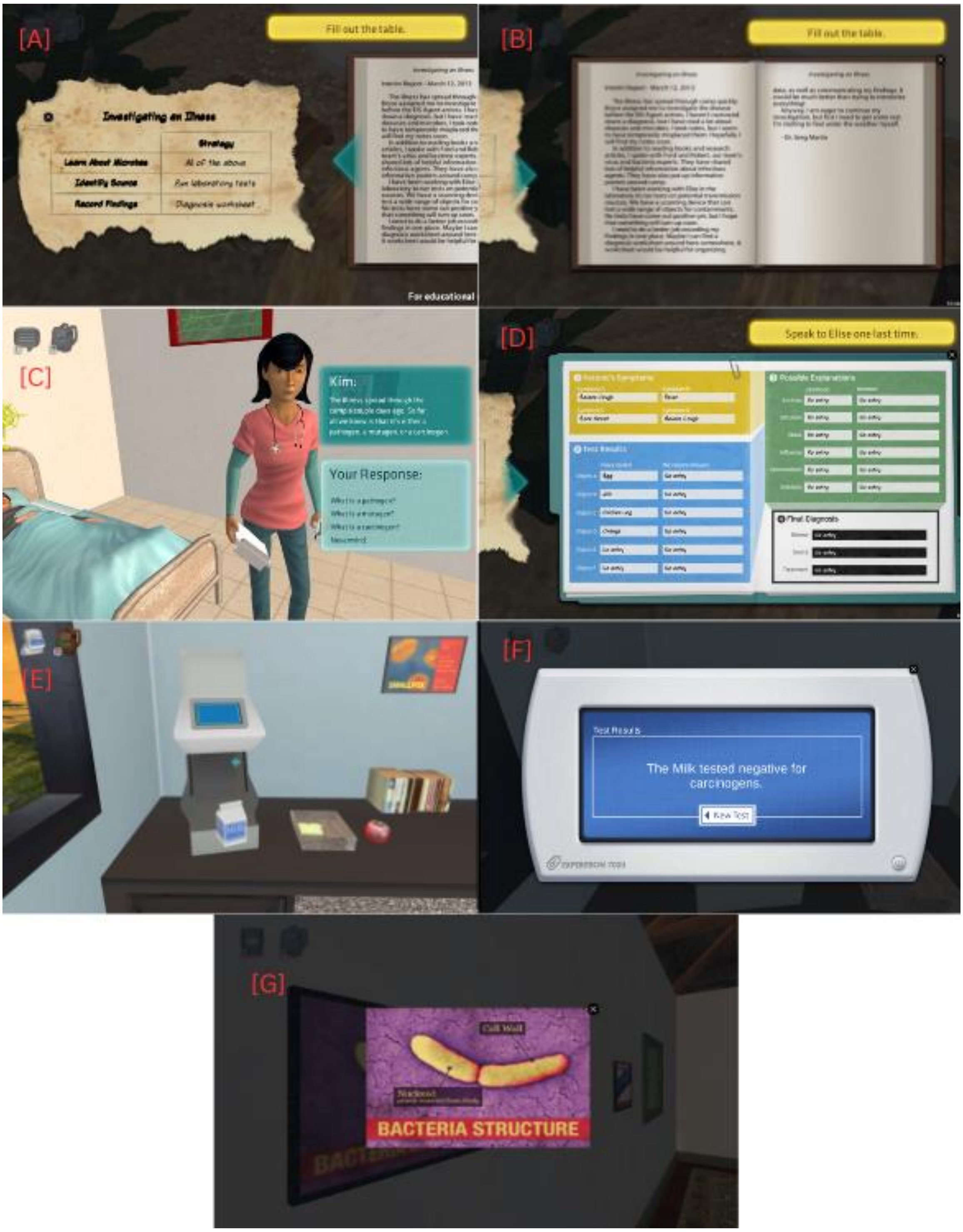

Figure 1. *Crystal Island* components. (A) Concept Matrix; (B) Book and Research Article; (C) NPC Interactions; (D) Diagnostic Worksheet; (E) Scanner (front view); (F) Scanner Results; (G) Poster

3.3 Coding & Scoring

3.3.1 Game Mechanics

Game Mechanics were defined using Plass et al.'s [6] Integrated Design Framework for Playful Learning as elements incorporated within a game to elicit learning by promoting learners' cognitive, affective, metacognitive, motivational, and social processes. This framework identifies learning and assessment mechanics as types of game mechanics which were contextualized to this study according to their operational definitions described in the literature review above.

For this study's purposes, we defined *learning mechanics* as tools embedded in the game that aid in participants' conceptual understanding of the domain (i.e., microbiology). Learning mechanics were further divided into subtypes including static, dynamic, and aid. Static learning mechanics included the books, research articles, and texts in which participants could open and close, but not directly change their interactions with the material. Dynamic learning mechanics included tools that provided domain information but in a way that was more customizable to the participants than the static subtypes. Specifically, dynamic learning mechanics included the conversations participants could have with NPCs in which participants could choose from several dialog options to obtain information about microbiology or the disease within the game. Aid learning mechanics were the tools provided to participants that consolidated knowledge without providing any feedback which, contextualized to *Crystal Island*, were the edits participants made to the diagnostic worksheet.

Assessment mechanics were defined as tools embedded within a game that evaluate participants' knowledge in relation to the developing domain knowledge and information that the participant is required to know to complete the game. As such, the assessment mechanic subtypes were identified as content assessment mechanics and game assessment mechanics. Content assessment mechanics were defined as the tools within the game which assessed participants' emerging understanding of microbiology. These mechanics contextualized to *Crystal Island* were concept matrices which revealed how much participants learned from the information provided within static learning mechanics (i.e., books and research articles). Game assessment mechanics were defined as the tools which assess participants' progress towards game completion and provided feedback to the participant on that progress. Within *Crystal Island*, these game assessment mechanics were both the scanner in which participants input the food item and hypothesized about the disease infecting researchers on the island and participants' final submission of the diagnosis. Both the scanner and worksheet submission evaluated how well the participant engaged in information gathering and scientific reasoning behaviors to complete the game.

3.3.2 Learning Outcomes

Two types of learning outcomes were evaluated: learning gains and game completion. *Learning Gains* is the difference between learners' pre-task and post-task microbiology quiz scores while accounting for prior knowledge and is calculated to identify the extent to which the *Crystal Island* environment increased learners' domain knowledge about microbiology. To calculate learning gains, we use a series of equations from Marx and Cummings [66].

Game Completion is defined as learners' ability to successfully solve the mystery within *Crystal Island*. Game completion is a binary factor in which a learner did solve the mystery (regardless of the number of attempts taken) by submitting a correct diagnosis, *TRUE* ($N = 62$), or a learner did not solve the mystery by either not submitting a diagnosis at all or incorrectly submitting one or more diagnoses, *FALSE* ($N = 75$).

3.4 Preliminary Analyses

To ensure that gender and was not a factor in participants' interactions with game mechanics, we conducted preliminary analyses to determine how these individual differences are related to learning outcomes. A chi-squared test revealed that game completion success did not vary by gender ($p > .05$) and a t-test revealed no significant differences in learning gains between genders ($p > .05$). As such, gender differences were not included in subsequent analyses.

Additionally, the data of interest in Research Questions 1-5 were frequencies which can be considered either count data or continuous data. To determine the type of data frequencies were considered within these analyses, we examined the ranges of learning mechanics, assessment mechanics and game mechanic subtypes (i.e., static, dynamic, aid, content, game). Further descriptives (i.e., skew and kurtosis) were assessed for normality of these variables to determine if non-parametric statistics were more appropriate compared to parametric statistics (see Table 1). From these statistics as well as further tests for homogeneity, we determined that parametric statistical tests (e.g., *t*-tests, ANOVAs) were appropriate for the analyses.

Table 1. Frequency, ranges, and skewness of game mechanic types and subtypes.

Frequency Type	Variable	Descriptives		Range		Skew & Kurtosis
		M	SD	Min	Max	
Game Mechanic Frequency	Learning Mechanics	54.5	15.5	17	113	<|2|
	Assessment Mechanics	37.7	15.6	9	78	<|2|
Game Mechanic Subtype Frequency	Static	21.9	9.76	0	63	<|2|
	Dynamic	11.6	4.27	2	28	<|2|
	Aid	11.7	6.75	0	31	<|2|
	Content	18.9	9.29	0	43	<|2|
	Game	18.5	11.3	0	51	<|2|

4. Results

4.1 Research Question 1: Do learning gains differ between groups of learners who successfully completed *Crystal Island* versus learners who did not successfully complete the game?

A *t*-test was used to identify if learners who successfully completed the game significantly differed in their learning gains compared to learners who did not successfully complete the game. Results showed no significant differences in the learning gains displayed by learners who successfully solved the game (M = 0.06; SD = 0.32) and those who did not (M = -0.02; SD = 0.36; p < .05). From this result, it was critical to understand how learners' use of game mechanics, in terms of frequencies, sequential transitions, and transitions across time on task, significantly related to how learners increased their science learning outcomes regarding both microbiology content learning and use of scientific reasoning to complete the game.

4.2 Research Question 2: Are learning outcomes related to the frequency in which learners deployed learning versus assessment game mechanics?

First, we wanted to see if there were differences between the frequency in deployment between learning mechanics and assessment mechanics by running a paired *t*-test. Results from this analysis showed that participants used significantly more learning mechanics (M = 54.5, SD = 15.5) than assessment mechanics (M = 37.7, SD = 15.6; $t(136)$ = 11.5, p < .01).

We then wanted to see if the frequency of game mechanics were significantly related to learning gains. Two correlations were run to identify if a relationship existed between learning gains and learners' frequencies of learning and assessment game mechanics. Results showed that while there was no significant relationship between learning mechanic frequency and learning gains (p > .05), there was a significant negative relationship between assessment frequency and learning gains ($r(135)$ = -0.20, p < .05). In other words, as participants engaged with increasingly more assessment mechanics, learning gains decreased.

Last, we ran two *t*-tests to identify if the frequency in which learning and assessment game mechanics were deployed differed between participants who were or were not successful in completing the game. Results showed that there was a significant difference in the frequency of learning mechanic use between participants who did and did not successfully complete the game ($t(122.7) = 3.34$, $p < .01$) where participants who successfully completed the game used learning mechanics significantly more ($M = 59.2$, $SD = 16.0$) than participants who did not successfully complete the game ($M = 50.6$, $SD = 14.1$). However, there were no significant differences in the frequency of assessment mechanics between participants who did ($M = 35.8$, $SD = 13.9$) and did not ($M = 39.9$, $SD = 16.8$) successfully complete the game ($p > .05$).

4.3 Research Question 3: Are we able to identify distinct clusters of learners based on their transition frequency across game mechanics?

To identify clusters of participants who were similar in their game mechanic use, we first identified the frequency in which participants transitioned across learning and assessment mechanics. For each participant, four frequencies were identified for each of the following sequential transitions: (1) learning to assessment; (2) learning to learning; (3) assessment to learning; and (4) assessment to assessment. We then classified clusters of participants using hierarchical clustering which is used to identify the participants whose data are most similar to each other, calculated through Euclidian distances. Figures 2 & 3 visually show how each participant was clustered due to their similarity in data. Figure 2 shows a dendrogram of each participant who is represented as an endpoint on the x-axis. Each participant is then connected through branches to other participants who show similar frequencies across all sequential transitions. The greater the height of the branch between participants, the greater the distance they are in terms of demonstrating similar data. Figure 3 circles the participants across the three final groups identified from our dendrogram which allows us to visually assess the variability between participants within the same cluster.

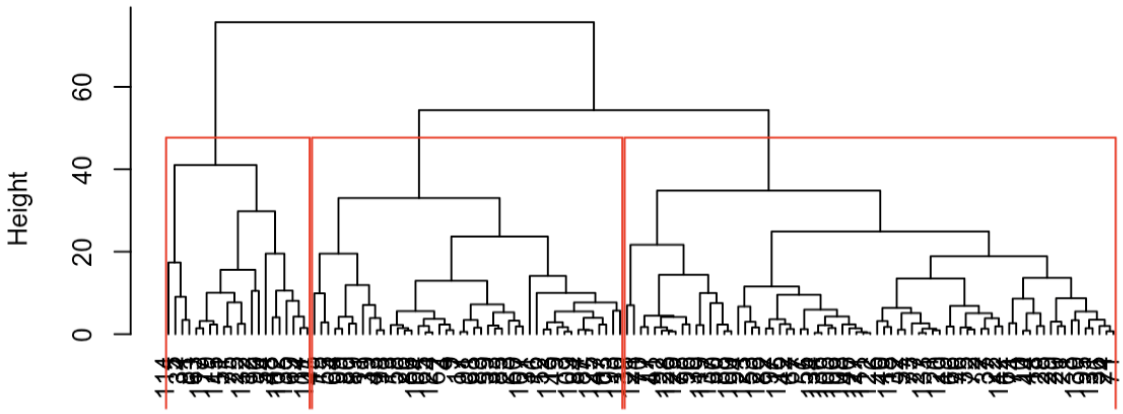

Figure 2. Dendrogram depicting participants and their similarity to each other. Red partitioning lines indicate the three clusters of participants used for further analyses.

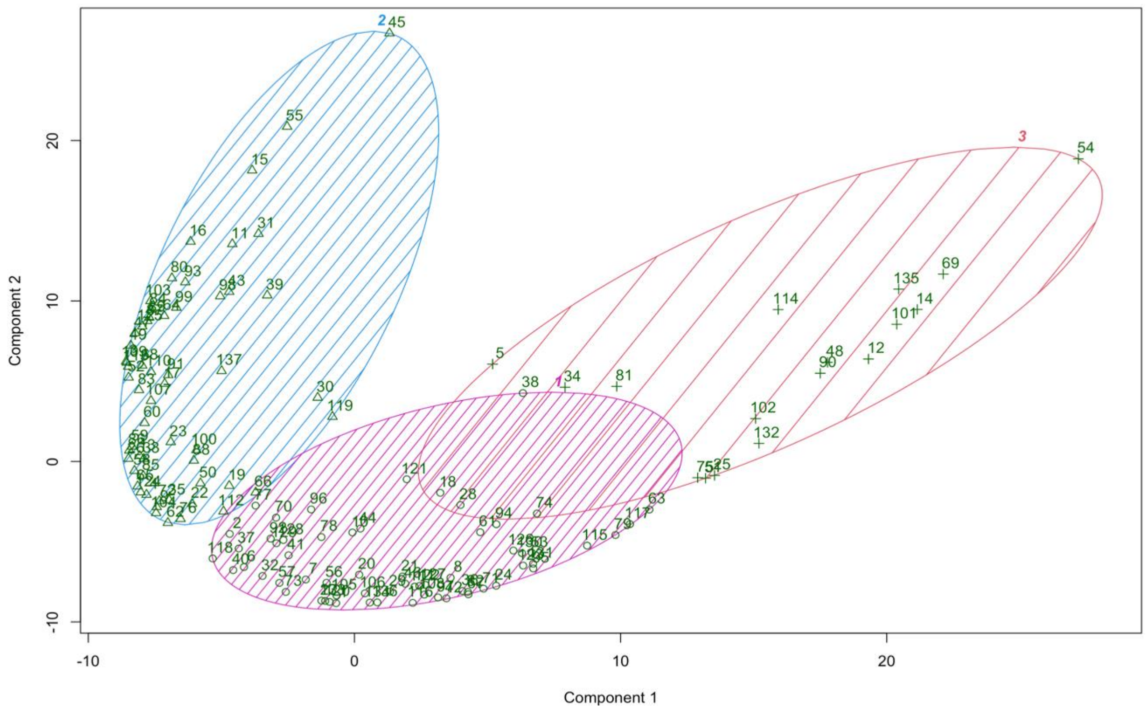

Figure 3. Visualization depicting how each participant was clustered within the three groups. Pink = Cluster 1, Blue = Cluster 2, Red = Cluster 3.

Three clusters of participants were identified using this technique. To identify the profiles of each cluster, a two-way mixed ANOVA was conducted to identify how each cluster differed depending on the frequency of their game mechanic transitions (see Table 2 for descriptive statistics). Results showed that there are significant main effects of cluster ($F(2, 134) = 5.58$, $p < .01$) and transition ($F(1.94, 260.1) = 334.0$, $p < .01$) with a significant two-way interaction effect between cluster and transition on frequency, $F(3.88, 260.1) = 60.6$, $p < .01$.

Table 2. Descriptive statistics for frequency of game mechanic transitions across each cluster.

Cluster	Learning > Assessment M(SD)	Learning > Learning M(SD)	Assessment > Learning M(SD)	Assessment > Assessment M(SD)
1 (N = 62)	17.5(4.8)	40.6(13.7)	16.8(4.8)	14.8(7.2)
2 (N = 58)	19.5(4.8)	30.0(9.33)	19.0(4.78)	29.1(11.0)
3 (N = 17)	8.59(2.98)	46.8(14.1)	7.88(2.91)	11.5(10.6)

Pairwise t-tests with a Bonferroni correction were then used to identify the differences between all clusters in the frequency of transitions across game mechanics. As there were twelve tests that were conducted, the alpha value for these comparisons is .004. Table 3 shows the statistics for these comparisons.

Table 3. Pairwise comparisons in the frequency of game mechanic transitions across clusters.

Transition	Cluster I	Cluster J	Mean Difference (I-J)	*p*-value (adjusted)
Learning > Assessment	1	2	-2.00	$p = .055$
	1	3	8.91	$p < .004$*
	2	3	10.9	$p < .004$*

Learning > Learning	1	2	10.6	*p* < .004*
	1	3	-6.2	*p* = 0.19
	2	3	-16.8	*p* < .004*
Assessment > Learning	1	2	-2.20	*p* = .02
	1	3	8.92	*p* < .004*
	2	3	11.1	*p* < .004*
Assessment > Assessment	1	2	-14.3	*p* < .004*
	1	3	3.3	*p* = .63
	2	3	17.6	*p* < .004*

Note. *denotes significance at α = .004, Bonferroni correction for alpha = 0.05 for 12 repeated tests.

To simplify the tables above, Cluster 1 is characterized by a high frequency transition from learning to assessment, learning to learning, and assessment to learning with low frequency transitions from assessment to assessment. Cluster 2 is characterized by a high frequency transition from learning to assessment, assessment to learning, and assessment to assessment with low frequency transitions from learning to learning. While Clusters 1 and 2 tended to have higher transition frequencies across a greater number of game mechanic transitions, Cluster 3 shows only a high transition frequency from learning to learning with all other transition frequencies being significantly lower than either Clusters 1 or 2. From these significance tests, we can conclude that the participants who belonged to Cluster 1 showed high frequency transitions but characteristically lower recurrent sequential transitions within assessment mechanics (i.e., assessment to assessment). Cluster 2 also showed high frequency transitions but characteristically lower recurrent sequential transitions within learning mechanics (i.e., learning to learning). Cluster 3 showed low frequency transitions but characteristically higher recurrent sequential transitions within learning mechanics, the opposite of Cluster 2.

4.4 Research Question 4: Do clusters of learners who differ in their frequency of game mechanic transitions vary in their learning gains and game success?

This research question aims to identify if a specific cluster of participants identified in the previous research question has greater learning gains or more participants who successfully completed the game. In identifying these differences between clusters, we can identify if participants' sequential transitions across learning and assessment mechanics are related to learning gains or game success. An ANOVA was first run to identify any differences in learning gains between clusters. Results revealed that there were no significant differences between clusters ($p > .05$), indicating that the process in which learning occurred within the cluster did not have a significant impact on learning gains measured by pre- and post-tests of microbiology knowledge.

A logistic regression was run to identify the probability participants successfully completed the game depending on the cluster they were assigned. From the results of this model, we found that cluster has a statistically significant effect on game success (Deviance = -9.07, $p < .05$). Results from this model show that participants in Cluster 2 are 60.5% less likely to successfully complete the game than participants in Cluster 1 (Std. Error = 0.38; z(134) = -2.43, $p < .05$). Within Cluster 2, 18 of 58 participants (31%) successfully completed the game whereas within Cluster 1, 33 of 62 participants (53%) successfully completed the game. However, the likelihood that participants in Cluster 1 will successfully complete the game compared to participants in Cluster 3 is relatively the same ($p > .05$) where within Cluster 3, 11 of 17 participants (65%) completed the game. As such, we continue this paper by further examining the differences between Clusters 1 and 2, excluding the 17 participants in Cluster 3, to better understand how these participants' transitions across learning and assessment game mechanics contribute to their success in completing the game.

4.5 Research Question 5: Is there a difference in the frequency in which learning and assessment game mechanic subtypes are deployed by clustered learners?

This research question aimed to determine if there were differences in the frequency in deployment across the subtypes of all game mechanics between Clusters 1 and 2. We used a two-way mixed ANOVA which used the subtype as the within-subjects variable and cluster, the between-subjects variable, and frequency of participants' use of these subtypes as the outcome. Results showed a non-significant main effect of cluster on the frequency of mechanic subtypes ($p > .05$) with a significant main effect of subtype ($F(2.02, 238.3) = 52.1$, $p < .01$). In other words, the frequency of participants' overall use of all game mechanics did not significantly differ between Clusters 1 and 2 but there were significant differences between the frequency of subtype use regardless of cluster. Additionally, there was a significant interaction effect between cluster and subtype ($F(2.02, 238.3) = 21.8$, $p < .01$) showing that the frequency in which participants used certain game mechanics differed depending on the cluster in which participants were grouped.

Pairwise *t*-tests with Bonferroni corrections were then conducted on each of the significant effects from the two-way mixed ANOVA. First, we examined paired pairwise comparisons between the frequency of different subtypes across all participants. We adjusted the p-value and set the alpha at the 0.005 level for the ten pairwise *t*-tests using a Bonferroni correction. Results found that, across all participants, static, content, and game subtypes were used more often than aid and dynamic subtypes (see Table 4).

Table 4. Pairwise comparisons for main effect of subtype on frequency.

Subtype	*M*	*SD*	1	2	3	4
Static	21.9	9.76	--	--	--	--
Dynamic	11.6	4.27	**-11.5*** [-12.2, 08.60]	--	--	--
Aid	11.7	6.75	**-11.6*** [-12.0, -8.51]	0.26 [-0.84, 1.09]	--	--
Content	18.9	9.29	-3.47 [-4.71, -1.29]	**7.67*** [5.48, 9.30]	**-6.70*** [-9.41, -5.12]	--
Game	18.5	11.3	**-2.55*** [-6.10, -0.76]	**-7.41*** [-8.82, -5.10]	-7.79 [-8.57, -5.10]	0.32 [-2.23, 3.09]

Note. * indicates significance at α = .005, Bonferroni correction of alpha = 0.05 for 10 repeated tests.

Second, we examined the pairwise comparisons of frequencies between both subtypes and Clusters 1 and 2. We adjusted the p-value with Bonferroni corrections for significance and set the alpha at the 0.01 level for the five pairwise *t*-tests. Results found that there were no significant differences between clusters in their frequency of dynamic learning mechanics but significant differences for all other subtypes. Participants in Cluster 1 had significantly greater frequencies of static and aid subtypes than those in Cluster 2 who had significantly greater frequencies of content and game subtypes (see Table 5).

Table 5. Pairwise comparisons of frequencies between mechanic subtypes and clusters.

Subtype	Cluster 1 M(SD)	Cluster 2 M(SD)	t-value
Static	24.2(10.8)	19.6(7.92)	2.65*
Dynamic	12.4(4.59)	10.7(3.76)	2.18
Aid	13.5(7.36)	9.74(5.45)	3.18*
Content	15.5(7.83)	22.6(9.36)	-4.53*
Game	14.6(9.21)	22.7(12.0)	-4.11*

Note. * indicates significance at α = .01.

4.6 Research Question 6: How do clusters of learners transition across subtypes of learning and assessment game mechanics?

To examine how participants within Clusters 1 and 2 transition across subtypes of learning and assessment mechanics, we used a First Order Markov Model to identify the probabilities in which participants transitioned from one subtype to another. Table 6 shows the probabilities of these transitions for both Clusters 1 and 2.

Table 6. Probabilities participants from Clusters 1 and 2 transitioned across learning and assessment mechanic subtypes.

Cluster 1					
From::To	Aid	Dynamic	Static	Content	Game
Aid	0.12	0.38	0.34	0	0.15
Dynamic	0.17	0.32	0.30	0	0.20
Static	0.15	0.08	0.32	0.43	0.02
Content	0.10	0.12	0.40	0.35	0.03
Game	0.28	0.07	0.06	0	0.55
Cluster 2					
From::To	Aid	Dynamic	Static	Content	Game
Aid	0.08	0.40	0.25	0	0.24
Dynamic	0.15	0.36	0.33	0	0.16
Static	0.08	0.07	0.23	0.61	0.02
Content	0.07	0.09	0.34	0.46	0.04
Game	0.19	0.04	0.04	0	0.71

We then visualized the differences between the cluster's probabilities of each transition within Figure 4. The edges (i.e., directional arrows connecting nodes) within the figure demonstrate the probability that the transition will occur across all participants. The nodes that appear represent the game mechanic subtype (i.e., dynamic, static, content, aid, game). The directionality of the arrows demonstrate the order in which the transition occurred. For example, the arrow leading from Dynamic into Static has a value of 0.3. This means that there is a 30% chance that participants who are engaging with a dynamic mechanic will subsequently transition into a static mechanic. Within this figure, only transitions that have at least a 10% chance of occurring are depicted. The darker the edge, the greater the probability that transition occurred. For example, the edge from Dynamic to Aid (value = 0.17) is significantly lighter than the edge from Aid to Dynamic (value = 0.38) because the transition from Dynamic to Aid is less likely to occur.

Edges were colored red when the probability of a participant from Cluster 1 performing that transition was at least 10% more likely than a participant from Cluster 2 performing that transition. Edges were colored purple when the probability of a participant from Cluster 2 performing that transition was at least 10% more likely than a participant from Cluster 1 performing that transition. For example, as the edge from Game to Aid is red, participants in Cluster 1 (value = 0.33) were more likely to engage in that transition compared to participants in Cluster 2 (value = 0.21). Further, as the edge from Static to Content is purple, participants from Cluster 1 (value =0.48) were less likely to engage in that transition compare to participants in Cluster 2 (value = 0.73). As such, this visualization has several functions including, visually identifying stronger probabilities of transitions, the transition directionality, and how Clusters compared to each other in terms of which transitions were more likely to occur depending on the cluster to which the participant belonged.

From this visualization we see that participants in Cluster 1 were more likely to transition: (1) from static to aid; (2) from content to aid; (3) from content to dynamic; (4) from aid to aid;

and (5) from game to aid. Conversely, participants in Cluster 2 were more likely to transition: (2) from static to content; (2) from content to content; (3) from aid to game; and (4) from game to game.

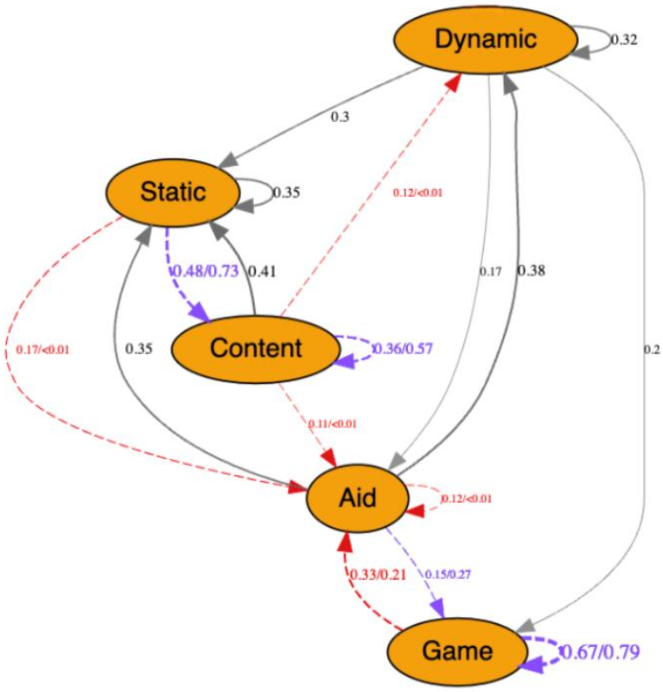

Figure 4. Probabilities that participants transitioned from one game mechanic subtype to another. Red edges denote probabilities for that transition is greater for Cluster 1 than Cluster 2. Purple edges denote probabilities for that transition is greater for Cluster 2 than Cluster 1.

.

5. Discussion

The goal of this study was to examine how game mechanics contributed to science learning. Specifically, this study wanted to understand how learners' transitions across learning and assessment game mechanics contributed to learning outcomes, situating learning as a process that is influenced by the design and pedagogical elements within educational games, not just as an outcome that can be assessed via standardized testing. The current study explored how high-school learners used game mechanics throughout gameplay and how this influenced learning gains and game success.

The first research question identified differences in learning gains between learners who were successful in completing game objectives and those who were unsuccessful. We hypothesized that learners who were able to complete the game would demonstrate greater scientific reasoning ability and therefore have greater learning gains. However, results did not support this hypothesis where there were no significant differences in learning gains between learners who were successful and unsuccessful. This result contradicted some of the prior literature in GBLEs (e.g., [19]) but also is supported by divergent previous work that found GBLEs are not always successful at supporting learning (e.g., [21],[33]-[35]). This finding helps motivate the rest of our paper to further examine learning game mechanics, which are used to increase domain knowledge, and assessment mechanics, which are used to quantify learners' progress towards obtaining game and learning objectives. By examining how learners interacted with these game mechanics differently, we can understand the contributions each type of game mechanic has for progressing the learning in achieving greater learning outcomes in terms of knowledge acquisition and demonstration of abilities.

From the findings of the first research question, our second research question examined if learning outcomes were related to the frequency in which learners deployed learning versus assessment game mechanics. We had two hypotheses related to each of the learning outcomes defined in this paper. First, we hypothesized that learners who had a greater frequency of learning and assessment mechanics would have greater learning outcomes. Second, we hypothesized that learners who successfully completed the game would demonstrate greater frequency of assessment mechanics.

Results from the second research question found that overall, learners tended to engage in significantly more learning mechanics than assessment mechanics. Further related to learning gains, while no relationships were found between learning gains and learning mechanic frequency, learners who had greater assessment mechanic frequency had lower learning gains. This contradicts prior literature where we would have expected to see greater learning gains from greater learning mechanic use due to the pedagogical grounding of learning mechanics [6]. As such, we see that the current learning mechanics, including large blocks of text, conversations with NPCs, and tools to synthesize information, are not significant accelerators of learning. This is not entirely surprising given the design conditions that previous work has cautioned against for trying to disguise traditional learning content under the guise of "games" (e.g., [32]). It is possible that these static learning mechanics within *Crystal Island* are too similar to traditional content to gain any benefits typically associated with GBLEs, as previous work has suggested [21].

Further, assessment mechanics were found to have a negative relationship to learning gains, contradicting prior literature as well [6],[37],[64]-[65] which would have stated that due to the assessment mechanic feedback provided to learners, learners should have been able to accurately gauge their learning. From this we see that learners, even with feedback on in-game assessments, are unable to engage in the self-regulatory skills needed to then re-visit important materials throughout the game. Additionally, this is also an indication that learners may have attempted to game the system through the increased use of assessment mechanics (which are necessary to complete the game) rather than learning about microbiology, leading to decreased learning gains. From these results, we recommend that games should provide greater amounts of scaffolding for directing learners to learning mechanics after the use of assessment mechanics with further constraints placed on the assessment mechanics used to end the game prior to learners' assessed domain knowledge.

Results from the second research question related to game completion found that learners who successfully completed the game engaged in significantly more learning mechanics with no significant difference in frequency of assessment mechanics compared to learners who did not successfully complete the game. This did not fully support our hypotheses in which we expected that learners who received feedback from assessment mechanics would have a greater chance of success [19], especially as assessment mechanics were required to be used for learners to attempt to complete the game. Across both research questions, results show that the pedagogical grounding of learning mechanics is imperative to game success rather than domain knowledge acquisition where learners are unable to effectively learn with assessment mechanics. Further studies are needed to fully understand why this effect occurs. Specifically, is this due to learners' shallow processing of domain information held within learning mechanics? Could learners be unable to identify relevant texts for learning versus those needed to complete the game? Do learners not find value in the feedback provided by in-game assessment mechanics? These questions could be answered in future research by collecting multimodal data that would capture the temporally unfolding cognitive and metacognitive processes within and across learning and assessment mechanics.

The third research question attempted to identify clusters of learners based off of learners' transition frequencies across game mechanics. For this research question as well as the subsequent research questions, the goal was to understand how learners' sequential transitions within and across learning and assessment mechanics were similar to other learners, if profiles of these learners could be identified, and if these profiles were related to greater learning outcomes. From previous studies [9], profiles of learners have been extracted in terms of the use of learning mechanics. Additionally prior studies have examined the transition across game mechanics [21]. As such, we hypothesized that we would be able to identify clusters of learners differing in their transition frequency across game mechanics. Results found three general clusters of learners. Cluster 1 was characterized by high frequency transitions but lower recurrent sequential transitions within assessment mechanics. Cluster 2 was characterized by high frequency transitions but lower recurrent sequential transitions within learning mechanics. Cluster 3 was characterized by low frequency transitions but higher recurrent sequential transitions within learning mechanics. From the identification and subsequent characterization of profiles of each cluster, further research questions were asked to identify the significance of these clusters in terms of learning outcomes. In doing so, we are able to build profiles of learners who may be more or less likely to be successful in terms of learning gains and game success, providing implications for future adaptivity of GBLEs. To further inform these feedback mechanism and adaptability suggestions, we then needed to examine the differences in learning outcomes across the clusters.

The fourth research question examines if there are differences in learning outcomes between clusters. We hypothesized that clusters of learners who varied in their game mechanic transition frequencies would demonstrate differences in learning outcomes where we may be able to identify profiles that demonstrate more optimal transitions over others. Results revealed that there were no significant differences in learning gains between clusters, indicating that the process in which learning occurred within the cluster did not have a significant impact on learning gains measured by pre- and post-tests of microbiology knowledge. However, we did find that Clusters 1 and 3 were significantly more likely to successfully complete the game than learners in Cluster 2. Due to the small sample size of Cluster 3 and their similarities to Cluster 1, Cluster 3 was excluded from analyses. As such, we conclude that learners who demonstrated lower recurrent sequential transitions within assessment mechanics are more likely to be successful in completing the game than learners with lower recurrent transitions within learning mechanics. As such, it is likely that encouraging learners to decrease their repeated use of assessment mechanics would lead to greater game success, i.e., successful demonstration of scientific reasoning abilities.

The fifth research question delved deeper into the differences between clusters in their learning and assessment subtypes. We initially hypothesized that learners across clusters would demonstrate differences in their game mechanic subtype use. Results partially supported this hypothesis where there were no differences between clusters on their overall game mechanic frequency, but there were differences in their use of game mechanic subtypes. Pairwise comparisons showed that learners in Cluster 1 had greater frequencies of static and aid subtypes (i.e., learning mechanics) whereas learners in Cluster 2 had greater frequencies of content and game subtypes (i.e., assessment mechanics) with no differences in the frequency of dynamic learning mechanic subtype use. This validates the profiles assigned to each cluster in which learners in Cluster 1 tended to have smaller recurrent use of assessment mechanics whereas Cluster 2 had smaller recurrent use of learning mechanics. Further, we see that static and aid learning mechanics are the driving mechanic subtypes which can assist learners in demonstrating scientific reasoning abilities, indicated by Cluster 1's higher probability of successful game completion. This extend prior literature on game mechanics [9],[33],[53],[63] and transforms how we view specific learning mechanics and their subtypes in which learning mechanics are not just for pedagogical support and to increase domain knowledge, but they are great supporters for learners' ability to demonstrate scientific reasoning abilities although they are not explicitly taught within these mechanics. These results have further implications for the use of NPCs as effective dynamic learning mechanic subtypes in which these NPCs were interacted with similarly across clusters of learners and as such deserve a greater number of studies examining the role conversational NPCs have in educational games.

The sixth research question visually examined how learners transitioned across subtypes of learning and assessment game mechanics given the clusters they were placed within. Supporting our hypothesis based on prior literature [21], we see that learners from different clusters engage in subtype transitions differently. Similar to previous results from this study, we see that learners in Cluster 1 were more likely to engage in transitions that related to learning mechanics rather than assessment mechanics whereas learners in Cluster 2 had significantly greater recursive transitions between game and content assessment mechanic subtypes. Cry to visually identify how learners from different clusters transition across these game mechanics to build even more robust and detailed profiles of learners for future adaptivity for science learning games.

Overall, our results indicate that game success is not depended upon learning gains and as such, these measures of learning outcomes and performance should be considered separately. Additionally, we are able to characterize and profile learners based on their behaviors in the GBLE where the transitions between different game mechanics are highly related to the ability for learners to demonstrate their scientific reasoning skills. Finally, we conclude that learning mechanics, rather than assessment mechanics, drive learners' success in game.

6. Limitations

This study and its defined mechanics are highly contextualized to *Crystal Island*. Although themes arise from the mechanics which are present within most educational games for science (e.g., *Operation ARA*), such as static learning mechanics or content assessment mechanics, this study does not represent the exhaustive list that other GBLEs may include. As this study's implications draws largely from the interaction and transition between these game mechanics, games that include other mechanics and mechanic subtypes should be aware of the limited generalizability.

Additionally, this study pulls only from log files without including other modalities for capturing learners' use and transitions across game mechanics. Specifically, the game mechanics captured with log files represent the absolute actions of the learner, not potential considerations. In other words, if a learner was considering using one game mechanic but never directly interacted with that game mechanic, this consideration and cognitive decision making was not accounted for within these analyses. Further, self-reports which may have revealed learners' intentions and expectations regarding the duality of learning objectives (i.e., learning about microbiology, completing the game) were not captured within these analyses.

Due to these shortcomings, this paper can be expanded upon by other studies including a greater number of mechanics used within GBLEs for science learning as well as studies which can capture interactions and cognitive processes that are not logged by the game using multimodal multichannel process data (e.g., eye tracking, concurrent verbalizations, facial expressions of emotions) with students of varying ages.

7. Future Directions

Future directions for learning and assessment mechanics in serious games offer exciting avenues for researchers and designers to explore and innovate. As the field continues to evolve, several vital areas emerge as focal points for advancement. Firstly, researchers can delve into developing more adaptive and personalized learning experiences within serious games [67]. By integrating artificial intelligence (AI) and machine learning techniques, games can dynamically adjust content, challenges, and feedback to tailor learning experiences to an individual's needs [68]. This personalized approach not only enhances engagement and effectiveness but also has the potential to revolutionize the learning experience for each player, tailoring the gameplay to their unique learning profile and significantly improving their learning outcomes, and scientific reasoning skills.

Additionally, integrating multimodal data analytics, which refers to the analysis of data from multiple sources such as eye-tracking, facial expressions, physiological responses, and concurrent verbalizations, presents an exciting frontier for advancing assessment mechanics in serious games [69]). Researchers can gain valuable insights into learners' cognitive and metacognitive processes, emotions, motivational states, and behaviors during gameplay. This rich data is not just informative but crucial in designing more effective assessment strategies that accurately measure learning outcomes and scientific reasoning and provide actionable feedback to learners.

In summary, future research directions for learning and assessment mechanics in serious games involve leveraging adaptive technologies, exploring immersive experiences, and harnessing multimodal data analytics. These areas are particularly important as serious games are increasingly being recognized as valuable tools in STEM education. However, there is still much to be explored and understood about how to optimize their use for learning and assessment. By focusing on these areas, researchers can drive innovation, improve learning outcomes, and unlock the full potential of serious games as practical educational tools.

Acknowledgments

This study was supported by funding from the National Science Foundation (DUE#1761178 and DRL#1661202) and the Social Sciences and Humanities Research Council of Canada (SSHRC 895-2011-1006). The authors would like to thank members of the SMART Lab at UCF and the IntelliMEDIA Group at North Carolina State University for their contributions.

Conflicts 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.

References

[1] National Assessment of Educational Progress (NAEP): NAEP Report Card: *Science*. (2019) https://www.nationsreportcard.gov/science/student-experiences/?grade=12

[2] L. A. Dardas, I. Khalaf, M. Nabolsi, O. Nassar, S. Halasa. "Developing an understanding of adolescents' knowledge, attitudes, and practices toward COVID-19," *The Journal of School Nursing,* Vol. 36, no. 6, pp. 430-441, 2020, doi:10.1177/1059840520957069 https://doi.org/10.1177/1059840520957069

[3] A. Erfani, R. Shahriarirad, K. Ranjbar, A. Mirahmadizadeh, M. Moghadami, "Knowledge, attitude and practice toward the novel coronavirus (COVID-19) outbreak: A population-based survey in Iran," *Journal of Helath Science and Surveillance Systems,* Vol. 11, no. 3, pp. 578-591, 2023, https://doi.org/10.2471/BLT.20.256651

[4] M. Hartt, H. Hosseini, and M. Mostafapour, "Game on: Exploring the effectiveness of game-based learning," *Planning Practice & Research*, vol. 35, pp. 589-604, 2020, doi: 10.1080/02697459.2020.1778859.

[5] Z. Y. Liu, Z. A. Shakih, F. Gazizova, "Using the concept of game-based learning in education," *International Journal of Emerging Technologies in Learning,* Vol. 15, no.15, pp. 53-64, 2020, doi: 10.3991/ijet.v15i14.14675.

[6] J. L. Plass, B. D. Homer, R. E. Mayer, C.K. Kinzer, "Theoretical foundations of game-based and playful learning," in *Handbook of Game-based Learning,* J. L. Plass, R. E. Mayer, B. D. Homer, Eds., The MIT Press, 2019, pp. 3-24.

[7] S. Vandercryusse, M. Vandewaetere, and G. Clarebout, "Game-based learning: A review on the effectiveness of educational games," in *Handbook of Research on Serious Games as Educational, Business and Research Tools*, M.M. Cruz-Cunha, Ed., IGI Global, 2012, pp. 628-647. https://doi.org/10.4018/978-1-4666-0149-9.ch032

[8] C. Wardoyo, Y.D. Satrio, D. Ma'ruf, "Effectiveness of game-based learning: Learning in modern education," *KnE Social Sciences*, vol. 4, no. 7, pp. 81-87, Apr. 2020, doi: 10.18502/kss.v4i7.6844.

[9] D. A. Dever, M. J. Amon, H. Vrzakova, M. D. Wiedbusch, E. B. Cloude, and R. Azevedo, "Capturing sequences of learners' self-regulatory interactions with instructional material during game-based learning using auto-recurrence quantification analysis," *Frontiers in Psychology*, vol. 13, 2022, doi: 10.3389/fpsyg.2022.813677.

[10] J. Rowe, L. Shores, B. Mott, J. Lester, "Integrating learning, problem solving, and engagement in narrative-centered learning environments," *International Journal of Artificial Intelligence in Education,* vol. 21, pp. 115-133, 2011, doi: 10.3233/JAI-2011-019

[11] M. Taub, R. Azevedo, A. E. Bradbury, G. C. Millar, and J. Lester, "Using sequence mining to reveal the efficiency in scientific reasoning during stem learning with a game-based learning environment," *Learning and Instruction*, vol. 54, pp. 93-103, Apr. 2018, doi: 10.1016/j.learninstruc.2017.08.005.

[12] M. Wiedbusch, D. Dever, A. Goslen, D. Carpenter, C. Marano, K. Smith, and R. Azevedo, (in press), "Contextualizing plans: Aligning students goals and plans during game-based inquiry science learning," *Proceedings of the 24th International Conference on Human-Computer Interaction*, Springer, Cham.

[13] D. F. Halpern, K. Millis, A. C. Graesser, H. Butler, C. Forsyth, and Z. Cai, "Operation ARA: A computerized learning game that teaches critical thinking and scientific reasoning," *Thinking Skills and Creativity*, vol.7, no. 2, pp. 93-100, 2012, doi: 10.1016/j.tsc.2012.03.006

[14] C. Galas, and D. J. Ketelhut, "River City, the MUVE," *Learning and Leading with Technology*, vol. 33, no. 7, pp. 31–32, 2006.

[15] B. Nelson, D. J. Ketelhut, J. Clarke, E. Dieterle, C. Dede, B. Erlandson, "Robust design strategies for scaling educational innovations: The River City MUVE case study," in *The Design and Use of Simulation Computer Games in Education*, B. E. Shelton, D. Wiley, Eds., Rotterdam, The Netherlands, Sense Publishers, 2007, pp. 209–231. https://doi.org/10.1163/9789087903121_013

[16] G. Gris, and C. Bengtson," Assessment measures in game-based learning research: A systematic review," *International Journal of Serious Games*, vol. 8, no. 1, pp. 3–26, 2021, doi: 10.17083/ijsg.v8i1.383.

[17] W. Westera, “Games are motivating, aren’t they? Disputing the arguments for digital game-based learning,” *International Journal of Serious Games*, vol. 2, no. 2 pp. 3-17, 2015. https://doi.org/10.17083/ijsg.v2i2.58
[18] P. Wouters, C. van Nimwegen, H. van Oostendrop, and E. van der Spek, “A meta-analysis of the cognitive and motivational effects of serious games,” *Journal of Educational Psychology,* vol. 105, no. 2, pp. 249-265, 2013, doi: 10.1037/a0031311.
[19] K. L. McClarty, A. Orr, P. M. Frey, R. P. Dolan, V. Vassileva, A. McVay, “A literature review of gaming in education,” Pearson, Jun. 2012.
[20] D. A. Dever, A. M. Banzon, N. A. M. Ballellos, R. Azevedo, “Capturing learners’ interactions with multimedia science content over time during game-based learning,” *Proceedings of the 1st Annual Conference of the International Society of the Learning Sciences*, p. 195-202, 2021.
[21] D. A. Dever, M. Wiedbusch, S. Park, A. Llinas, J. Lester, and R. Azevedo, “Assessing the complexity of game mechanic use during science learning,” in Proc. 12th GALA., G. Goos and J. Hartmanis, Eds., Nov. 2023, pp. 299-308. https://doi.org/10.1007/978-3-031-49065-1_29
[22] S. E. Huber, A. Lindstedt, K. Kiili, Ninaus, M. Ninaus, “Are game elements fueling learners’ motivation via positive affect?,” in proceedings of the 12th annual games and learning alliance conference., P. Dondio,M. Rocha, A. Brennan, A. Schönbohm, F. de Rosa, A. Koskinen, F. Bellotti, Ed. Springer Nature Switzerland, 2023, p. 234-243. https://doi.org/10.1007/978-3-031-49065-1_23
[23] L. Zhong, “Investigating learning efficiency and mental efficiency in a personalized role-playing-game environment,” *Interactive Learning Environments*, pp. 1–12, Dec. 2022, doi: 10.1080/10494820.2022.2157840.
[24] S. Arnab, T. Lim, M. B Carvalho, F. Bellotti, S. De Freitas, S. Louchart, N. Suttie, R. Berta, and A. De Gloria, “Mapping learning and game mechanics for serious games analysis,” *British Journal of Educational Technology*, vol. 46, no. 2, pp. 391-411, 2015, doi: 10.1111/bjett.12113.
[25] J. L. Plass, B. D. Homer, C. K. Kinzer, “Foundations of game-based learning,” *Educational Psychologist,* vol. 50, no. 4, 258-283, 2015, doi: 10.1080/00461520.2015.1122533
[26] K. Isbister, M. Flanagan, C. Hash, “Designing games for learning: Insights from conversations with designers,” *Proceedings of the 2010 CHI conference on human factors in computing systems*, p. 2041-2044, 2010, doi: 10.1145/1753326.1753637.
[27] S. Pawar, F. Tam, J. L. Plass, "Emerging design factors in game-based learning: Emotional design, musical score, and game mechanics design,” in *Handbook of Game-based Learning,* J. L. Plass, R. E. Mayer, B. D. Homer, Eds., The MIT Press, 2019, pp. 347-365.
[28] P. H. Winne, “Cognition and metacognition within self-regulated learning,” in *Handbook of Self-Regulation of Learning and Performance* (Educational Psychology Handbook Series). D.H. Schunk and J.A. Greene, Eds., 2nd ed. New York, NY, USA.: Routledge/Taylor & Francis Group, 2018, pp. 36–48. https://doi.org/10.4324/9781315697048-3
[29] B. DiSalvo,” Pink boxes and chocolate-dipped broccoli: Bad game design providing justifications for reluctant learners,” in Proc. 11th GLS., A. Ochsner, eds. Jul. 2015.

[30] C. Harteveld, "Triadic game design: Balancing reality, meaning and play," *Springer Science & Business Media*, 2011. https://doi.org/10.1007/978-1-84996-157-8
[31] V. J. Shute, F. Ke, R. G. Almond, S. Rahimi, G. Smith, X. Lu, "How to increase learning while not decreasing the fun in educational games," In *Feldman*, R, Ed., McGraw Hill, 2019, pp. 327-257.
[32] J. Kellinger, "Repurposing tech tools for game-based learning" in *Smart Pedagogy of Game-based Learning*, L. Daniela, Ed., Switzerland: springer, 2021, pp. 41-55. https://doi.org/10.1007/978-3-030-76986-4_3
[33] D. A. Dever, R. Azevedo, E. B. Cloude, and M. Wiedbusch, "The impact of autonomy and types of informational text presentations in game-based environments on learning: Converging multi-channel processes data and learning outcomes," *International Journal of Artificial Intelligence in Education*, vol. 30, pp. 581-615, 2020, doi: 10.1007/s40593-020-00215-1.
[34] W.S. Ravyse, S. Blignaut, V. Leendertz, A. Woolner, "Success factors for serious games to enhance learning: A systematic review," *Virtual Reality,* vol. 21, pp. 31-58, 2017, doi: 10.1007/s10055-016-0298-4
[35] A. Tlili, M. Chang, J. Moon, Z. Liu, D. Burgos, and N. S. Chen, "A systematic literature review of empirical studies on learning analytics in educational games," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 7, no. 2, pp. 250-261, 2021, doi: 10.9781/ijimai.2021.03.003.
[36] C. L. Brown, M. A. Comunale, B. Widgahl, and S. Urdaneta-Hartmann, "Current climate for digital game-based learning of science in further and higher education," *FEMS Microbiology Letters,* vol. 365, no. 21, 2018, doi: 10.1093/femsle/fny237
[37] S. Rahimi, J. T. Walker, L. Lin-Lipsmeyer, J. Shin, "Toward defining and assessing creativity in sandbox games," *Creativity Research Journal,* vol. 36, pp. 1-19, 2023, doi: 10.1080/10400419.2022.2156477
[38] K. E. DiCerbo, Y. Xu, R. Levy, E. Lai, and L. Holland," Modeling student cognition in digital and nondigital assessment environments," *Educational Assessment*, vol. 22, no. 4, pp. 275-297, 2017, doi: 10.1080/10627197.2017.1382343.
[39] F. Quansah, "Traditional or performance assessment: What is the right way to assessing learners," *Research on Humanities and Social Sciences,* vol. 8, no. 1, pp. 21-24, 2018
[40] V. J. Shute, X. Lu, S. Rahimi, G. Smith, F. Ke, R. Almond, C. P. Dai, R. Kuba, Z. Liu, X. Yang, and C. Sun, "Maximizing learning without sacrificing the fun: Stealth assessment, adaptivity and learning supports in educational games," *Journal of Computer Assisted Learning*, vol. 37, no. 1, 2021, pp. 127-141, doi: 10.1111/jcal.12473.
[41] A. Emerson, E. Cloude, R. Azevedo, and J. Lester, "Multimodal learning analytics for game-based learning," *British Journal of Educational Technology*, vol. 51, no. 5, pp. 1505-1526, 2020, doi: 10.1111/bjet.12992.
[42] R. Azevedo and D. Gašević, "Analyzing multimodal multichannel data about self-regulated learning with advanced learning technologies: Issues and challenges," *Computers in Human Behavior*, vol. 96, pp. 207-210, 2019, doi: 10.1016/j.chb.2019.03.025
[43] M. S. El-Nasr, T. H. D. Nguyen, A. Canossa, and A. Drachen, "Game data science," *Oxford University Press*, 2021. https://doi.org/10.1093/oso/9780192897879.001.0001
[44] M. Taub, R. Azevedo, A. Bradbury, G. C. Millar, J. Lester, "Using sequence mining to reveal the efficiency in scientific reasoning during STEM learning with a game-based

learning environment," *Learning and Instructions,* vol. 54, p. 93-103, 2018. https://doi.org/10.1016/j.learninstruc.2017.08.005
[45] Y. Pan, F. Ke, C.-P. Dai, "Patterns of Using Multimodal External Representations in Digital Game-based Learning," *Journal of Educational Computing Research,* vol. 60, no. 8, pp. 1918-1941, 2023, doi:10.1177/07356331221087771
[46] R. Azevedo, F. Bouchet, M. Duffy, J. Harley, M. Taub, G. Trevors, E. Cloude, D. Dever, M. Wiedbusch, F. Wortha, and R. Cerezo, "Lessons learned and future directions of MetaTutor: Leveraging multichannel data to scaffold self-regulated learning with an intelligent tutoring system," *Frontiers in Psychology*, vol. 13, 2022, doi:10.3389/fpsyg.2022.813632.
[47] R. Sawyer, J. Rowe, J. Lester, "Balancing learning and engagement in game-based learning environments with multi-objective reinforcement learning," *Proceedings of the 18th international conference on artificial intelligence in education,* pp 323-334, 2017, doi: 10.1007/978-3-319-61425-0_27
[48] D. Rajendran, S. Prasanna, "Stealth assessment strategy in distributed systems using optimal deep learning with game based learning," *The Journal of Supercomputing,* vol. 78, pp. 8285-8301, 2022, doi: 10.1007/s11227-021-04236-y
[49] W. Min, M. H. Frankosky, B. W. Mott, J. P. Rowe, A. Smith, E. Wiebe, ... J. Lester, "DeepStealth: Game-based learning stealth assessment with deep neural networks," IEEE Transactions on Learning Technologies, vol.13, no. 2, pp. 213-325, 2020, doi: 10.1109/TLT.2019.2922356
[50] J. Frommel, C. Phillips, and R. L. Mandryk, "Gathering self-report data in games through NPC dialogues: Effects on data quality, data quantity, player experience, and information intimacy," Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, p. 1-12, 2021, doi: 10.1145/3411764.3445411.
[51] V. J. Shute, X. Lu, and S. Rahimi, "Stealth assessment," in *The Routledge Encyclopedia of Education*, J. M. Spector, Ed., London, UK.: Taylor & Francis group, 2021, pp. 1-9. https://doi.org/10.4324/9781138609877-REE58-1
[52] G. Smith, V. J. Shute, S. Rahimi, C. P. Dai, and R. Kuba, "Stealth assessment and digital learning game design," in *Games as Stealth Assessment*, M. P. McCreery and K. Kach, Eds., Hershey, PA, USA.: IGI Global, 2023, pp. 81-100. https://doi.org/10.4018/979-8-3693-0568-3.ch004
[53] M. Taub, N. Mudrick, R. Azevedo, G. Millar, J. Rowe, and J. Lester, "Using multi-channel data with multi-level modeling to assess in-game, performance during gameplay with CRYSTAL ISLAND," *Computers in Human Behavior*, vol. 76, pp. 641-655, Nov. 2017, doi: 10.1016/j.chb.2017.01.038.
[54] R. Mayer, "Educational psychology's past and future contributions to the science of learning, science of instruction, and science of assessment," *Journal of Educational Psychology,* vol. 110, no. 2, pp.174–179, 2018, doi:10.1037/edu0000195.
[55] E. B. Cloude, D. A. Dever, M. D. Wiedbusch, and R. Azevedo," Quantifying scientific thinking using multichannel data with crystal island: Implications for individualized game-learning analytics," *Frontiers in Education*, vol. 5, 2020, doi: 10..3389/feduc.2020.572546.
[56] M. J. Gomez, J. A. Ruipérez-Valiente, P. A. Martinez, and Y. J. Kim, "Exploring the affordances of sequence mining in educational games," in Proc. 8th TEEM., F J. García-Peñalvo and A. García-Holgado, eds. Oct. 2020. pp. 648-654. https://doi.org/10.1145/3434780.3436562

[57] R. Sawyer, J. Rowe, R. Azevedo, J. Lester, “Filtered Time Series Analyses of Student Problem-solving Behaviors in Game-based Learning,” *Proceedings of the 11th International Conference on Educational Data Mining,* p. 229-238, 2018.
[58] D. Carpenter, E. Cloude, J. Rowe, R. Azevedo, and J. Lester, “Investigating student reflection during game-based learning in middle grades science,” *Proceedings of the 11th International Learning Analytics and Knowledge Conference*, p. 280-291, 2021, doi: 10.1145/3448139.3448166.
[59] A. Emerson, E. Cloude, R. Azevedo, and J. Lester, “Multimodal learning analytics for game-based learning,” *British Journal of Educational Technology*, vol. 51, no. 5, pp. 1505-1526, 2020, doi: 10.1111/bjet.12992.
[60] M. Geden, A. Emerson, D. Carpenter, J. Rowe, R. Azevedo, and J. Lester, “Predictive student modeling in game-based learning environments with word embedding representations of reflection,” *International Journal of Artificial Intelligence in Education*, vol. 31, pp. 1-23, 2021, 10.1007/s40593-020-00220-4.
[61] A. Gupta, D. Carpenter, W. Min, J. Rowe, R. Azevedo, and J. Lester,” Detecting and mitigating encoded bias in deep learning-based stealth assessment models for reflection-enriched game-based learning environments*,” International Journal of Artificial Intelligence in Education*, 2023, doi: 10.1007/s40593-023-00379-6.
[62] M. Taub, R. Sawyer, A. Smith, J. Rowe, R. Azevedo, and J. Lester, “The agency effect: The Impact of student agency on learning, emotions, and problem-solving behaviors in a game-based learning environment,” *Computers & Education*, vol. 147, 2020, Art. No. 103781, doi: 10.1016/j.compedu.2019.103781.
[63] A. Emerson, R. Sawyer, R. Azevedo, J. Lester, "Gaze-enhanced student modeling for game-based learning," *Proceedings of the 26th conference on user modeling, adaptation and personalization*, p. 63-72, 2018. https://doi.org/10.1145/3209219.3209238
[64] Y. C. Liu, W. T. Wang, and T. L. Lee, “An integrated view of information feedback, game quality, and autonomous motivation for evaluating game-based learning effectiveness,” *Journal of Educational Computing Research,* Vol. 59, no. 1, pp. 3-40, 2021, doi:10.1177/0735633120952044.
[65] P. H. Winne, R. Azevedo, “Metacognition and self-regulated learning,” in *The Cambridge Handbook of the Learning Sciences*, R. Sawyer, Ed., Cambridge University Press, 2022, pp. 93–113, doi: 10.1017/97811 08888295.007
[66] J.D. Marx and K. Cummings, “Normalized change,” *American Journal of Physics,* Vol. 75, no. 1, pp. 87-91, 2007, doi: 10.1119/1.2372468.
[67] J. L. Plass, S. Pawar, “Adaptivity and personalization in game-based learning,” in *Handbook of Game-based Learning*, J. L. Plass, R. E. Mayer, B. D. Homer, Eds., The MIT Press, 2020, pp. 263-281.
[68] J. Lester, M. Bansal, G. Biswas, C. Hmelo-Silver, J. Roschelle, J. Rowe, “The AI Institute for Engaged Learning,”*AI Magazine*, Vol. 45, no. 1, pp. 69-76, 2024. https://doi.org/10.1002/aaai.12161
[69] M. Giannakos, D. Spikol, D. Di Mitri, K. Sharma, X. Ochoa, R. Hammad, “Introduction to Multimodal Learning Analytics,” in *The Multimodal Learning Analytics Handbook,* M. Giannakos, D. Spikol, D. Di Mitri, K. Sharma, X. Ochoa, R. Hammad, Eds., Springer, 2022, pp. 3-28, doi: 10.1007/978-3-031-08076-0_1