International Journal of Artificial Intelligence in Education (2020) 30:581–615
https://doi.org/10.1007/s40593-020-00215-1

ARTICLE

The Impact of Autonomy and Types of Informational Text Presentations in Game-Based Environments on Learning: Converging Multi-Channel Processes Data and Learning Outcomes

Daryn A. Dever, et al. *[full author details at the end of the article]*

Accepted: 30 August 2020 / Published online: 15 September 2020

Abstract
Game-based learning environments (GBLEs) focus on enhancing learning by providing learners with various representations of information (e.g., text, diagrams, etc.) while allowing full autonomy, or control over their actions. Challenges arise as research shows that learners inaccurately use cognitive and metacognitive processes when given full autonomy. This study examined 105 undergraduates who were randomly assigned to autonomy conditions (i.e., full, partial, and no autonomy) as they interacted with scientific informational text presentations (i.e., non-player characters [NPCs], books and research articles, posters) during learning with Crystal Island, a GBLE. We assessed how learners' eye-tracking (e.g., fixation durations on objects) and log-file (e.g., durations of activities) data reflected how learners interacted with text presentations and selected pretest-relevant items (i.e., text providing answers to questions on the pretest). Results showed that participants in the partial autonomy condition ($n = 38$) demonstrated higher learning gains than those in the full autonomy condition ($n = 45$). Time spent interacting with all books and research articles within Crystal Island were positively correlated with learning gains. There were significant differences in learners' duration and fixation duration on informational text presentation interactions between conditions and within types of presentations as well as significant interactions between pretest-relevant items and types of presentations. Overall, autonomy and pretest relevancy impact the time interacting with informational text presentations which influence learning. Implications are provided for applying autonomy during game-based learning, and how this may direct future implementations of AI within GBLEs to provide implicit scaffolding via adaptively limiting learners' autonomy as they interact with informational text.

Keywords Autonomy · Game-based learning environment · Eye tracking · Log files · Scaffolding

[1]An earlier version of this paper was presented at the AIED 2019 conference (Chicago, IL) and published as Dever & Azevedo (2019a).

Introduction

Game-based learning environments (GBLEs) afford autonomy while giving learners access to a myriad of game elements including books, posters, non-player characters (NPCs), scientific data, etc. with the implicit assumption that more autonomy, or learners' ability to control their own actions during learning (Bandura 2001), leads to better learning and problem solving (Plass et al. 2015). In addition, narrative-centered GBLEs contain storylines that support informational content (e.g., scientific text) and provide real-world scenarios ranging across multiple domains for learners to practice problem-solving, critical thinking, cognitive, and metacognitive skills (Chen et al. 2018; Rowe et al. 2009; Shih et al. 2015). Thus, the role of narrative within GBLEs is critical in conveying informational content to enhance learning. Information within GBLEs include multiple representations of instructional materials (e.g., graphics, pictures, videos, text) designed using multimedia learning principles with aims of improving learning (Mayer 2014; Plass et al. 2015). While the majority of research on GBLEs tend to focus on learning outcomes and using self-reports to measure motivation, engagement, and so forth (see Plass et al. 2020), we examined the role of autonomy and types of informational text presentations by converging college students' learning outcomes and process data (e.g., eye movements and log-files) during learning with a GBLE designed to foster learning about microbiology.

Affording autonomy in GBLEs increases opportunities to engage learners by allowing them to freely utilize authentic instructional materials and interact with contextualized problem-solving scenarios that require scientific reasoning and self-regulatory processes (e.g., solving a mystery related to a biological outbreak on an island). These interactive scenarios aid in transforming complex, rich information into easily interpretable concepts that are contextualized to engage and support learners in acquiring knowledge (Chen et al. 2018; Herrington et al. 2014). Although learners within GBLEs are able to engage in exploring various instructional materials through problem-solving activities (Lee et al. 2011) by affording autonomy, effective learning with GBLEs requires learners to accurately and dynamically monitor and regulate their cognitive, affective, metacognitive, and motivational processes, such as distinguishing between relevant and irrelevant instructional materials, evaluating competing hypotheses, comprehending informational texts, changing plans, and so forth (Azevedo et al. 2018; Rowe et al. 2009). Typically, when provided with autonomy in a GBLE, learners tend to inaccurately monitor and regulate their self-regulatory processes by engaging in off-task behaviors as they are unable to deal with extraneous cognitive load, fail to comprehend relevant instructional materials, etc., thus leading to small or negligible learning gains and not completing the objectives of the GBLE (Rowe et al. 2009).

Crystal Island (Rowe et al. 2011) is a narrative-centered GBLE designed to support learning about microbiology while fostering scientific reasoning and self-regulated learning skills. It accomplishes this by providing learners with varying levels of autonomy and providing a range of game elements and tools to solve the mysterious illness on the island. By including informational content (i.e., scientific books and research articles, concept matrices measuring reading comprehension, NPCs including a camp nurse, patients, and scientists, scientific worksheet to record symptoms and hypotheses, scientific tools to analyze microbes), Crystal Island assists learners in obtaining content knowledge in microbiology and skills in scientific reasoning while self-regulating. Data generated from learners while interacting with these elements (e.g., durations on scientific text) could provide insight into their cognitive and metacognitive skills as they learn with a GBLE; however, most studies using GBLEs fail to capture and synchronize multiple channels of data (e.g., eye-tracking, log-file) to capture learning through these interactions.

Taub et al. (2016) investigated metacognition, specifically strategy use, while participants learned with Crystal Island by examining fixation durations on book and research article content as well as concept matrices to predict performance which was measured by the number of concept matrix attempts. This study found that, by using eye-tracking data, researchers can identify how learners strategize when collecting information from a book or research article to predict learners' in-game performance based on the number of books and research articles a learner uses and their proportion of fixations on the content (Taub et al. 2016). In another study, Emerson et al. (2018) used eye tracking to develop a framework to predict learners' performance and cognition with Crystal Island. This study emphasized that learners' eye-tracking data on in-game elements significantly improve models predicting problem-solving performance. Emerson et al. (2018) used multiple actions to measure performance, including which NPCs learners interacted with and the amount of time spent interacting with each NPC, testing a correct or incorrect object, the number of attempts on in-game measures (e.g., concept matrices), the efficiency of the learners' actions in relation to solving the problem, and the number of solution attempts.

Another study by Dever and Azevedo (2019a) examined whether varying levels of autonomy impacted how learners interacted with NPCs, books, and research articles. Their results showed that learners with partial autonomy had higher proportional learning gains and longer fixation durations on books, research articles, and NPC interactions than learners with full autonomy, highlighting that autonomy influences amount of time spent interacting with NPCs, books, and research articles and subsequent learning outcomes with GBLEs. Taub et al. (2018) examined how metacognition affected game efficiency and completion using log files. They found performance was measured by identifying how often learners' tested relevant, partially-relevant, and irrelevant objects. Learners who were defined as being more efficient in solving the game had significantly fewer instances of testing partially-relevant and irrelevant items than learners who were less efficient. Similarly, Dever and Azevedo (2019b) examined learners' selection of relevant from irrelevant information that would enable the learner to, not only solve the final diagnosis but, perform well on the domain pre- and posttests. These were examined using eye-gaze data, where fixation duration proportions between different goal-directed actions (e.g., scanning items, completing concept matrices) were compared and reading books and research articles and completing their concept matrices were identified as the greatest contributors to learners' overall time in game. As such, this study identified how learners' fixation duration and identification of the relevancy of the books and research articles affected their learning. Further, this study by Dever and Azevedo (2019b) identified that learners who spent a greater time revisiting relevant books and research articles had significantly greater learning gains than those who spent more time revisiting irrelevant books and research articles. This shows that learners identifying the relevancy of information within text presentations is critical for learning.

In contrast to the previous studies that used one type of process data (i.e., eye gaze), Taub et al. (2017) combined eye tracking and log files to investigate cognitive and metacognitive self-regulated learning during in-game performance measures. Their findings showed that learners with low proportions of fixations on information and the in-game performance measure demonstrated higher learning outcomes than those with high proportions of fixations on information and the in-game performance measure. These studies emphasize the using process data generated during learning with a GBLE has the potential to effectively accurately capture learners interacting with elements in the game, and how these variables are related to performance. However, the aforementioned studies do not directly investigate *how* learners interact with and learn using scientific information given external constraints (i.e., autonomy) imposed by a GBLE. Further, these studies demonstrate gaps where process data could indicate how learners use metacognitive strategies to select, organize, and integrate scientific information to enhance their learning using multiple data channels.

The current study differs from previous studies in multiple ways to address this gap in the literature. First, the current study addresses all sources of information within Crystal Island, where NPCs, books and research articles, as well as posters were defined as sources of information integral to accurately completing and performing well on the domain pre- and posttests. Secondly, this study extends previous work by Dever and Azevedo (2019a, b) and is in contrast to the other previous studies, where performance is measured by participants' learning gains to examine how learners' domain knowledge changes as a function of their interactions with different types of text within Crystal Island. Lastly, the current study identified the relevancy of information within several types of text presentations in reference to the items on the pre-test. By doing so, the current study adopted assumptions that all information contained in the pre- and posttests are addressed within the GBLE, not all information in the GBLE are relevant to the domain pre- and posttests, multiple sources of information may be identified as relevant to the domain pre- and posttests, and what is relevant to the final diagnosis, or the GBLE itself, may not be relevant to the domain pre- and posttests. This study 1) addressed how autonomy influences metacognition as captured through learners' process data while they interact with scientific information within and 2) identified gaps in current literature and explore the relationship between these two components (autonomy and interaction with scientific information) specifically within GBLEs.

Autonomy as Scaffolding in Game-Based Learning

GBLEs are designed to integrate autonomy as scaffolding techniques used to increase engagement and learning. More specifically, levels of autonomy have been implicitly designed in GBLEs to facilitate learning, increase motivation, and so forth, but have not been explicitly tested as a scaffold that supports learning, comprehension, problem solving and reasoning. Providing autonomy combined with learners' inability to use cognitive and metacognitive strategies within GBLEs (Azevedo and Hadwin 2005; Sabourin et al. 2013; Taub & Azevedo, 2018) pose a challenge to learning. Previous research identified scaffolds as ways to guide learners within learning environments to support their understanding of complex information by providing adaptive scaffolding to foster learners' knowledge and skill acquisition, where the presence of these scaffolds fade as learners gain competence (e.g., Azevedo and Hadwin 2005; Pea 2004; Plass et al. 2015). Researchers have used different types of scaffolds including static, dynamic, conceptual, procedural, metacognitive, reflection, etc. with various technologies, especially intelligent tutoring systems (ITSs) such as AutoTutor, Affective AutoTutor, Betty's Brain, Gaze Tutor, MetaTutor, where they are delivered by pedagogical agents that prompt strategy and tool use, such as note-taking of concepts after reading to increase learning outcomes (Azevedo et al. 2018; Winne and Hadwin 2013; D'Mello et al. 2012; Graesser et al. 1999). However, while these prompts have been used extensively in ITSs and other learning technologies, they have not been widely used or empirically tested in GBLEs.

Implicit scaffolds (e.g., limited autonomy) unobtrusively support learners' knowledge and skill acquisition but rely on learners' competent, timely, and accurate use of cognitive and metacognitive processes without interrupting the learning experience with GBLEs. Studies have used autonomy, such as the amount of control learners have over how they interact during game-based learning, and assume learners, or agents, *actively* engage with elements when they select, organize, and integrate information presented in these environments such as informational content presented through video, text, pictures (Bandura 2001). Autonomy allows learners to make their own choices and initiate planning behaviors, but for some learners it comes with negative consequences such as lack of skill acquisition, minute learning gains, and so forth because of both internal (e.g., low prior knowledge, lack of self-regulatory skills) and external conditions (e.g., full autonomy in a GBLEs without explicit scaffolding). Specific to GBLEs, learners must constantly choose which game elements or tools to use and their course of action to achieve the objectives for completing the game (Bandura 2001). Varying levels of autonomy determine the control learners have over their choices and actions in the environment, where full autonomy gives learners complete control over planning, generating learning goals, decisions, use of tools, etc. In contrast, restricting autonomy reduces the amount of control learners have over their planning, generating learning goals, decisions, use of tools, etc. From a self-regulatory perspective, full autonomy in GBLEs is ideal if learners are capable of accurately and dynamically monitoring and regulating their cognitive, affective, metacognitive, and motivational processes, whereas more restrictive forms of autonomy are optimal for learners have challenges accurately and dynamically monitoring and regulating these process (Azevedo and Hadwin 2005; Mayer 2019; Sabourin et al. 2013; Taub & Azevedo, 2018). Limiting autonomy (i.e., learners ability to select, organize, and integrate information) in GBLEs enhances learning by using an implicitly fixed, procedural scaffold. Yet, the tradeoff of less autonomy is less engagement during learning (Sabourin et al. 2013; Plass et al. 2013, 2015) such that full autonomy encourages engagement but poses a threat to learning when learners cannot accurately apply cognitive and metacognitive strategies in selecting relevant informational content (Sabourin et al. 2013). The current study explores limited autonomy as a scaffold to support learners in selecting relevant information from scientific texts in a GBLE.

Theoretical Framework: Cognitive Theory of Multimedia Learning

GBLEs, such as Crystal Island, include multiple types and representations of information such as books and posters that learners must read and comprehend by selecting, organizing, and integrating relevant information in order to learn about microbiology and therefore solve the mystery. We used Mayer's (2014) Cognitive Theory of Multimedia Learning (CTML) which is based on three basic assumptions: 1) there are two separate processing channels that learners use to gather and interpret visual (e.g., picture) and verbal (e.g., audio) information; 2) learners are limited in the amount of information they can process simultaneously within each channel; and 3) learners are active in processing given information (e.g., Burkett & Azevedo, 2012; Butcher 2014). In addition, CTML addresses five cognitive processes which are combined into three phases: (1) selecting, (2) organizing, and (3) integrating information. *Selecting* refers to identifying relevant information from multimedia such as text and images. *Organizing* involves developing cognitive models from information selected, thus prompting the *integration* of prior knowledge into learners' new cognitive models to create an updated model of knowledge.

While not explicitly related to autonomy, self-regulation, and GBLEs, we extended Mayer's (2014) CTML by arguing that the second and third assumptions closely align with previous assumptions about autonomy where learners actively process visual and verbal information presented as multimedia instructional materials, requiring learners to dynamically and accurately select, organize, and integrate multiple representations of information embedded in GBLEs (such as Crystal Island) by using their monitoring and self-regulatory skills as needed, depending on the amount of autonomy. In summary, we used and extended Mayer's CTML by testing the impact of varying levels of autonomy (in versions of the same GBLE), on learners' interactions with informational text and resulting learning gains.

Application of Metacognitive Processes, Autonomy, and CTML

CTML integrates cognitive processes and, when contextualized to learning, addresses learners' metacognitive processes and competencies. For instance, according to CTML, a learner must first select relevant information from informational content presented to learners within non-dynamic learning environments (e.g., tutoring systems, multimedia). Relevant information is identified as content critical to learners' goals. For example, the goal of Crystal Island is to learn microbiology concepts by interacting with the environment. As such, goals that pertain to learning domain content may be defined by the presence of an explicitly communicated goal (e.g., instructions on how to complete a GBLE), or through covert methods (e.g., pretest items detailing concepts learned within a GBLE). Within GBLEs, learners must constantly monitor their progress towards goals as they select, organize, and integrate information they identified as relevant to the overall objective (e.g., learning domain knowledge; Azevedo et al. 2018). As such, CTML indirectly incorporates learners' metacognitive processing, where learners select information relevant to their goal (Greene and Azevedo 2009). Selecting relevant informational content is difficult to achieve for learners who are given little direction in deciding which information to select that will aid in achieving goals to increase learning and complete the game (Greene et al. 2010).

Therefore, in this paper, we argue that the level of autonomy is an integral part of learners' ability to successfully demonstrate metacognitive processes.

Previous literature has primarily used CTML to study non-dynamic environments (e.g. Cierniak et al. 2009). However, some studies have used CTML with applications to military training (e.g., Serge 2014) and foreign language acquisition (e.g., Alghamdi 2016) within dynamic learning environments (e.g., GBLEs, augmented reality). Limited research has examined how CTML can be applied to GBLEs that contain informational content in STEM domains and how, in combination with external constraints (i.e., autonomy), these factors influence cognitive and metacognitive processes and their relation to learning. Further, the limited number of studies examining CTML with GBLEs (e.g., Serge 2014) have not incorporated various process data necessary for examining cognitive and metacognitive processes. Therefore, there is a major gap in using CTML to study GBLEs, and how this model contributes to measuring learning, examining learners' interactions with game elements, and using process data to capture and understand cognitive and metacognitive processes critical for learning with GBLEs.

Process Data in Game-Based Learning

Multichannel process data facilitates researchers' ability to infer the cognitive and metacognitive processes that learners engage in by capturing and analyzing learners' interactions with learning environments (Azevedo and Gasevic 2019). This term, used throughout metacognition and self-regulated learning literature, refers to the variables which originate from multiple data streams (e.g., log files, eye tracking; Azevedo and Taub 2020). This paper emphasizes the use of multi-channel process data, log files and eye tracking, to examine how learners interact with multiple text presentations within a GBLE.

Log Files during Game-Based Learning

A plethora of studies have harnessed and analyzed in-game behaviors using log files since these data capture the frequency and duration at which learners initiate actions during game-based learning. A study by Taub et al. (2018) investigated whether log files could distinguish between scientific-reasoning and problem-solving behaviors during game-based learning. By utilizing sequential pattern-mining analysis, log files revealed two distinct groups where participants were efficient and less efficient in their scientific-reasoning behaviors related to completing the game. Another study by Cheng et al. (2015) examined log files and their relation to conceptual learning and game performance. They found that the frequency and duration of viewing relevant information was associated with game performance, where the more frequent and longer time spent viewing relevant information were positively associated with game performance and conceptual learning (Cheng et al. 2015). Similarly, Spires et al. (2011) examined middle-school students learning outcomes based on their scientific-reasoning actions (i.e., hypothesis vs. experimental actions) using log files. Results indicated that generating effective hypotheses during problem solving was positively associated with higher learning outcomes and game performance. Recently, studies have introduced models to assess students' developing knowledge and skills based on their in-game actions using log files (Shute 2011). A study by Akram et al. (2018) proposed a temporal-analytics framework that uses recurrent neural networks, a class of deep-learning methods that account for the temporal sequences in learners' log files, to analyze problem-solving strategies. Specifically, this analytical framework clustered students into groups based on the sequence of problem-solving strategies during game-based learning to develop predictive models that gauged competency and performance. From these studies, log files are primarily used to assess and predict learners' cognitive processes, learning, and performance as they interact with GBLEs.

Major challenges persist as researchers solely rely on log files to quantify learning and infer cognitive and metacognitive self-regulated learning processes (Azevedo et al. 2018; Winne 2018). Log files provide time-stamps for all learners' in-game actions, but do not provide fine-grained contextual information such as which elements learners were looking at when (e.g., content) they opened (i.e., they provide info that a specific book in a GBLE was open for a certain amount of time and preceded by another action and subsequently led to another action). We argue that log files need to be supplemented with finer-grained information supplied from eye-tracking data to examine attention allocation, gaze behaviors, and other relevant information that can be used to infer cognitive and metacognitive processes.

Eye-Tracking Methodology during Game-Based Learning

A large portion of studies have used eye tracking to investigate learning with GBLEs as it has been shown to reveal implicit indices of intent, reasoning, cognition, metacognitive monitoring, and decision-making processes (Taub et al. in press; Chen and Tsai 2015; Lai et al. 2013; Park et al. 2016). A study by Kiili et al. (2014) investigated pre- and post-test performance and its relation to total fixation and saccade duration on relevant and irrelevant informational content as learners interacted with GBLEs. They found that patterns in eye-gaze behaviors indicated when learners did not identify content relevant to the learning objective (Kiili et al. 2014). Tsai et al. (2016) utilized eye tracking to assess differences in eye-gaze behaviors between high- and low-domain knowledge groups as they completed a problem-based learning task using a GBLE. They found that participants who had little to no domain knowledge before learning with the game demonstrated longer and more frequent fixations on most game features in the GBLE compared to the high prior knowledge group. This study suggests that differences in eye-tracking data to mental exertion when taking prior knowledge into account, further supporting when learners demonstrate cognitive processing required to select and organize new information and integrate the new information into a coherent model. Another study investigated eye-gaze patterns as participants solved various problems with differing levels of difficulty (Lin 2014). This study concluded that longer time spent fixating on difficult problems relative to less difficult problems was indicative of higher cognitive load due to complex processing required for more difficult problems. These studies show that eye-tracking allow researchers to detect, measure, and understand cognitive processes related to selecting, organizing, and integrating information, and to examine how these processes relate to learning outcomes, performance, and comprehension during game-based learning (Mayer 2019; Plass et al. 2020). However, major challenges continue to persist because, while eye-tracking captures where learners allocate their attention and fixate on elements during learning with GBLEs, it fails to directly capture learners' level of understanding. For instance, if learners fixate on text in a book within a GBLE, eye-gaze behavior can indicate which information learners selectively attend to, but not the extent to which information was understood. A study by O'Keefe et al. (2014) used fixation durations and transitions between areas of interest to examine how multiple representations within a science simulation corresponded to learning with high school students. While this study found that fixation durations on multiple representations were not related to learning, the eye-gaze transitions between multiple representations can indicate learners' comprehension and transfer of illustrated concepts (O'Keefe et al. 2014). This study emphasizes the limitation of using fixation durations in connection to learning as well as the strength of using gaze transitions between elements within a learning environment to examine learning processes. However, the aforementioned studies use either log files or eye tracking to provide evidence of learning and performance. By converging multichannel data, our paper mitigates limitations and provides evidence of overt and covert cognitive and metacognitive processes that learners' employ while completing a single action within GBLEs.

Combining Eye-Tracking and Log Files during Game-Based Learning In this current study, we harnessed the strengths of both eye tracking and log files as two critical data channels in examining underlying cognitive and metacognitive processes during learning about microbiology with Crystal Island. Empirical studies show that capturing multiple channels of process data to classify learning processes is superior than a single data channel (Alonso-Fernández et al. 2019; Di Mitri et al. 2019; Giannakos et al. 2019). Specifically, a study by Taub et al. (2017) assessed cognitive and metacognitive self-regulatory processes using eye-gaze and log-file data with a GBLE, and how these in-game behaviors related to learning and performance. Their findings highlighted that combining eye-gaze and log-file data during game-based learning taps into the quality of cognitive and metacognitive processes such that log files capture the quantity of in-game actions, but eye-gaze data reveal more information on the quality of cognitive and metacognitive processing. Similarly, a study by Dever and Azevedo (2019a) examined eye-gaze and log-file data to investigate metacognition and how its use related to textual comprehension and performance during game-based learning. Their results showed that eye-gaze and log-file data quantified, not only when a participant was opening textual information, but also fixating on it and suggested higher learning gains were predicted by both the frequency and duration learners spent examining informational texts. However, critical gaps exist as few studies have compared and combined eye-tracking and log-file data to quantify and understand learning processes involved in game-based learning, and how they contribute to comprehension and learning.

Current Study

To address the major gaps in literature related to evaluating scaffolding in GBLEs and using CTML to assess learners' metacognitive competency in selecting relevant information when accounting for autonomy, this study examined how learners interacted with various types of informational text presentations, selected relevant information, and whether these interactions differed between levels of autonomy. We further examined how these constructs impacted learning captured using multichannel process data (i.e., eye tracking, log files) during learning with a GBLE, Crystal Island. Within this paper we addressed four research questions:

> Research Question 1) Do prior knowledge and learning gains significantly differ between learners with varying levels of autonomy?
> Research Question 2) Do learners' process data for each type of informational text presentation predict learning gains?
> Research Question 3) Do learners' varying levels of autonomy influence how learners interact with each type of informational text presentation?
> Research Question 4) Do learners' varying levels of autonomy and the relevancy of informational text influence how learners interact with each type of presentation?

We directly address these research questions by examining multichannel process data generated during learning with Crystal Island to capture how learners interacted with different types of informational text presentations and whether these interactions vary based on level of autonomy, the relevancy of the informational text content, and whether these process data are related to learning. For the *first research question*, we hypothesized that participants' pretest scores measuring prior knowledge of microbiology will not differ between conditions due to the randomization of the assigned autonomy conditions. Further, we hypothesized that normalized change scores will be higher for those in the partial agency condition than the full agency condition, as learners who are unable to accurately demonstrate cognitive and metacognitive processes have greater learning gains when control over their actions is restricted (Sabourin et al. 2013). For the *second research question*, we hypothesized that types of informational text presentations containing text and diagrams (e.g., NPCs, posters) will positively predict normalized change scores, as CTML states that learning with text and diagrams will increase learning compared to one informational presentation. For the *third research question*, we first hypothesized that participant interactions captured using process data will differ across text presentations, where participants will spend more time on rich scientific text provided by books, or books that were text-dependent and provided more information on a single, complex topic. Secondly, we hypothesized that participants in the partial agency condition will have greater durations on informational text than those in the full agency condition. Lastly, for the *fourth research question*, we hypothesized that participants in the partial agency condition will demonstrate greater durations and fixation durations on informational text presentations across pretest-relevant presentations than those in the full agency condition. This is supported by previous literature (Mayer 2014; Sabourin et al. 2013) as learners who are provided scaffolding will demonstrate a greater ability to identify and select relevant information compared to those with no support.

Methods

Participants and Materials

A sample of 120 undergraduate students were recruited from a large North American public university and participated in this study to learn about microbiology with a GBLE. However, only 105[2] undergraduate students (68.7% female), split between three conditions, full agency ($N = 48$), partial agency ($N = 35$), and no agency ($N = 32$) were included in our analyses due to missing data points, and measurement errors (e.g., eye-tracking calibration errors). The no agency condition is included only in specific research questions due to the nature of the condition itself, the data that is available for the condition, and the nature of the research questions. Ages ranged from 18 to 29 ($M = 20.0$, $SD = 1.80$).

Upon written consent, participants were administered a range of questionnaires before learning with the GBLE, which included demographics questions to gauge age, gender, ethnicity, and familiarity with video games (e.g., type of game, weekly length of play time) and self-report questionnaires to capture participants' emotions and motivation. We do not provide additional information about these scales to maintain concision in this paper as these data were not included in our statistical models. After participants answered the demographics and self-report items, they completed a 21-item, 4-option multiple-choice microbiology pre-test developed by a domain expert. These items addressed a range of topics such as the shape of a cell to identifying a genetic disease given a list of symptoms. All information measured on the pretest was provided in Crystal Island through various informational sources such as dialogue with NPCs (see Fig. 1) and reading books (see Fig. 2) and posters (see Fig. 3). After participants finished the objectives of the game, they were immediately administered another set of questionnaires to gauge self-efficacy for learning science, emotions, and motivation. Additionally, we administered a similar 21-item, 4-option microbiology posttest to capture knowledge gained after learners interacted with the GBLE. We excluded one pre- and posttest item from our analyses since we operationalized actions based on their relevance to pre-test items and one of the corresponding pretest items has conflicting information provided in the game. For example, participants could gather information about the reproduction of bacteria and viruses, but while one book provides the correct answer (i.e., bacteria can produce sexually or asexually), a poster and book both state a different answer (i.e., bacteria reproduce asexually). The question, when referring to the correct information regarding viral and bacterial reproduction, does not provide a correct answer. Therefore, this question was excluded from both pre- and posttests and a total score out of 20 was modified for these analyses.

Crystal Island Environment

Crystal Island (Rowe et al. 2011), a narrative-centered GBLE, encourages learners to develop scientific-reasoning and problem-solving skills along with learning microbiology. Learners within this environment are charged with diagnosing fellow researchers

[1] In the original AIED 2019 paper (i.e., Dever and Azevedo 2019a) we reported on 90 participants originating from the same study.

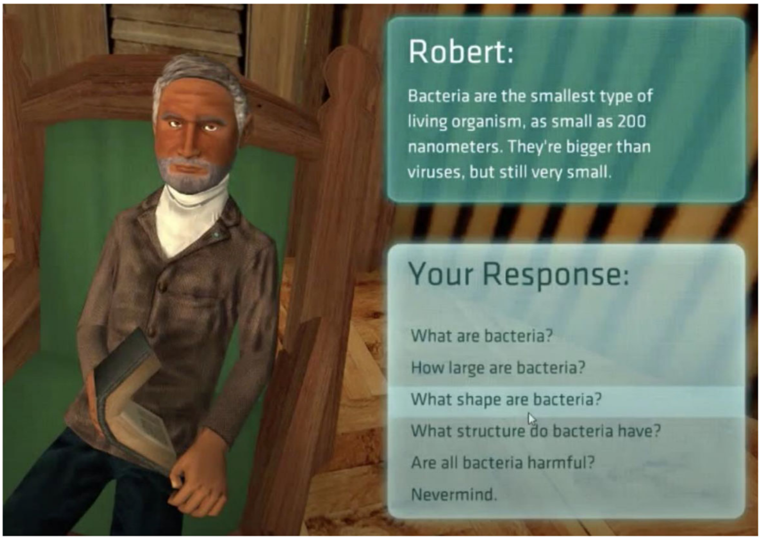
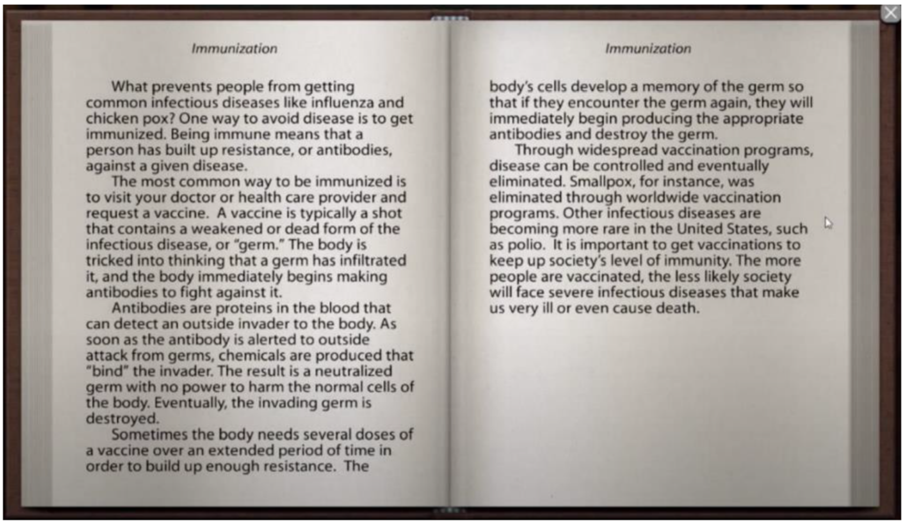


Fig. 1 Example of an NPC in Crystal Island

on a remote island who have contracted a mysterious illness by interacting with different elements (see Fig. 2) provided by the environment. To solve the mystery, learners must provide the name (e.g., influenza, salmonellosis), source (e.g., bread, milk, eggs), and treatment (e.g., rest, vaccination) of a disease to complete the game. Once on the island, learners converse with non-player characters (NPCs) who provide information for either solving the mystery (e.g., symptoms) or information related to domain content knowledge, (e.g., the size of bacteria). Learners are also provided informational content in the form of books and research articles scattered around the island which contain blocks of text related to domain knowledge and may be used to


Immunization

What prevents people from getting common infectious diseases like influenza and chicken pox? One way to avoid disease is to get immunized. Being immune means that a person has built up resistance, or antibodies, against a given disease.

The most common way to be immunized is to visit your doctor or health care provider and request a vaccine. A vaccine is typically a shot that contains a weakened or dead form of the infectious disease, or "germ." The body is tricked into thinking that a germ has infiltrated it, and the body immediately begins making antibodies to fight against it.

Antibodies are proteins in the blood that can detect an outside invader to the body. As soon as the antibody is alerted to outside attack from germs, chemicals are produced that "bind" the invader. The result is a neutralized germ with no power to harm the normal cells of the body. Eventually, the invading germ is destroyed.

Sometimes the body needs several doses of a vaccine over an extended period of time in order to build up enough resistance. The

Immunization

body's cells develop a memory of the germ so that if they encounter the germ again, they will immediately begin producing the appropriate antibodies and destroy the germ.

Through widespread vaccination programs, disease can be controlled and eventually eliminated. Smallpox, for instance, was eliminated through worldwide vaccination programs. Other infectious diseases are becoming more rare in the United States, such as polio. It is important to get vaccinations to keep up society's level of immunity. The more people are vaccinated, the less likely society will face severe infectious diseases that make us very ill or even cause death.

Fig. 2 Example of a book in Crystal Island

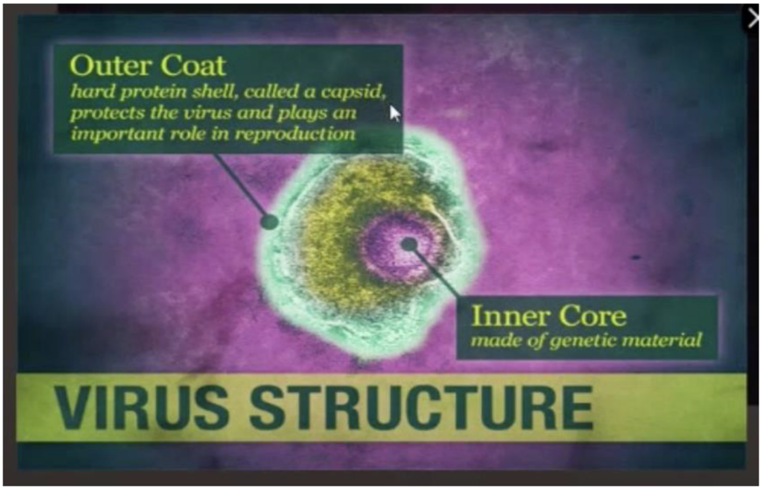


Fig. 3 Example of a poster in Crystal Island

complete the game. These books and research articles contain concept matrices which are used as performance measures, testing on the information from the corresponding text. Posters within Crystal Island provide short, sometimes uninformative, text and visuals that may coincide with the mystery of Crystal Island or the domain knowledge. For example, a poster may show an example of bacteria or virus structures with labels indicating the location of these structures. Often, the posters redundantly overlap information presented in books and research articles or conveyed by NPCs. Other tools and elements in Crystal Island include a scanner to hypothesize the disease and test food items to see if that disease is present. Learners are also given a diagnostic worksheet which allows for the documentation of symptoms, likeliness of the correct diagnosis to be a certain disease, and results from the scanning process. All items are necessary to complete the game.

Experimental Conditions

Participants were randomly assigned to one of three conditions after giving informed consent. Each of the conditions differed in level of autonomy participants were afforded during learning with Crystal Island. Specifically, the (1) *full agency* condition allocated complete autonomy to participants such that they could initiate any actions without constraints during learning. These actions included selecting elements to interact with such as books and research articles and testing food items when the participants wished, whereas the (2) *partial agency* condition set constraints on participants' actions by establishing a "ideal" path that participants were required to follow in order to complete the game. For example, participants were required to visit buildings in a specific sequence aimed at maximizing informational content acquisition (e.g., going to a building with information about influenza and then another building to talk with a non-player character about the symptoms of influenza). In the (3) *no agency* condition, participants watched a playthrough of the game using a third-person perspective without interacting with the game elements such as books and research articles or control the playthrough video (e.g., play, pause). This restricted any autonomy as participants learned with Crystal Island. These conditions were developed to represent the varying levels of autonomy that may be present throughout a GBLE to implicitly scaffold learners as they interact with the environment. As with most GBLEs, the full agency condition within Crystal Island represents the state of most GBLEs and how learners typically interact with these learning environments. Within this condition, participants are not provided any scaffolding while selecting and reading informational text presentations. Conversely, the partial agency condition implicitly scaffolds learners as they directly interact with the environment. The no agency condition did not allocate any level of autonomy to the participants as they learned with Crystal Island, and conversely did not serve as an implicit scaffold or directly support the participant. As such, we did not include participants in the no agency condition in parts of our analyses, but we include information about the conditions for replicability purposes.

Between the full and partial agency conditions, there were differences in their time on task. On average, participants in the full agency condition completed the game within 80.47 ($SD = 19.97$) minutes, while those assigned to the partial agency condition completed the game in 93.74 ($SD = 15.71$) minutes. While participants within different conditions differed in their time on task, they were not constrained in the time they could spend within their environment. In addition, total time in game accounts for multiple actions including editing the worksheet, completing concept matrices, and scanning food items. As such, total time on task was not considered as a covariate for the analyses, but rather the proportion of time spent on informational text presentations and the frequency of interactions learners had with NPCs, books and research articles, and posters (see Preliminary Analyses).

Apparatus

In this study, we captured eye-gaze behaviors and log files of each participant. An SMI RED250 eye tracker was used to collect participants' eye gaze behavior. Participants were calibrated using a 9-point calibration. This eye tracker was screen-based which sat at the bottom of the computer screen and had a sampling rate of 250 Hz which recorded 250 samples per second. Data from this eye tracker provided fixations, fixation durations, and regressions which where based off of predetermined areas of interest (AOIs). Log files were captured using information from the mouse and keyboard. This included the selection, or mouse clicks, of certain elements and objects in the Crystal Island environment, the time spent within one element (i.e., duration), and the movement of the participant throughout Crystal Island.

Participants' facial expressions of emotion and electrodermal activity (EDA) were identified using facial recognition of emotion software, implemented using iMotions FACET software run through Attention Tool 6.3 (iMotions 2016), and a physiological bracelet respectively. Data from the facial recognition and EDA bracelet were collected using Attention Tool 6.3 and analyzed using the baseline from each participant. The emotion recognition software analyzed 10 different emotions (e.g., happiness, sadness, confusion, anger, etc.) as participants interacted with Crystal Island. The EDA bracelet collected participants' skin conductance and heart rate. These data were collected, but not used for the analyses in the current study as the interpretation of these data and their units of measures were out of the paper's scope.

Experimental Procedure

Upon entering the laboratory setting, a researcher greeted participants and instructed them to sit in front of a complete to complete a series of questionnaires such as demographics questions and multiple self-report scales as well as a microbiology pretest to capture prior knowledge of the domain. Self-report scales included the Achievement Emotions Questionnaire (Pekrun et al. 2011) and the Achievement Goals Questionnaire (Elliot and Murayama 2008). Participants spent approximately 35 min to complete the pretest measures. Once the measures were completed, participants were calibrated to the SMI EYERED 250 eye tracker using a 9-point calibration to accurately measure individual eye-gaze behaviors. Next, participants were instructed to express a neutral facial expression and remain calm during calibration to the facial recognition of emotions software and electrodermal bracelet to determine a baseline that were captured using the Attention Tool 6.3 (see Fig. 4 of participant set up). After successful calibration, participants started learning with Crystal Island. Upon starting the game, participants were told that to complete the game, they needed to accurately diagnose the mystery illness plaguing the research camp and provide an appropriate solution treat the disease (e.g., influenza). The importance of searching for clues using the various tools (e.g., books, research articles, conversing with non-player characters) provided to participants during game-based learning were emphasized during the tutorial phase of the game. As participants interacted with elements in Crystal Island, we captured their process data that ranged from eye movements (e.g., fixation and saccades), facial expressions of emotions (e.g., neutral, joy, frustration), and log files (e.g., time spent engaging in actions). After participants completed the game by providing the correct treatment solution to the mystery illness, they were administered a posttest to measure differences in microbiology knowledge. After participants finished the task, we administered several self-report questionnaires which addressed different concepts (i.e., motivation, interest) than those administered for the pretest. Posttests included the Intrinsic Motivation Inventory (Ryan 1982), Perceived Interest Questionnaire (Schraw et al. 1995), and the Presence Questionnaire (Witmer and Singer 1998). Collectively, participants completed posttest measures in approximately 35 min. The pre- and posttest self-report scales were not considered as this study did not examine emotions, motivation, or self-efficacy, but directly questioned how process data measured through eye tracking and log files are utilized in understanding participant interactions with game elements. Afterwards, the researcher debriefed, compensated \$10/h (up to \$30), and thanked the participants for their time.

Coding and Scoring

Types of Informational Text Presentations

Crystal Island presents information in three different ways: (1) informative text with an uninformative visual (i.e., the visual does not add additional information that is not already provided in the text such as text on influenza with a picture of red blood cells), (2) informative text with no visuals (e.g., books and research articles), and (3) interchangeably informative and uninformative text and visuals (e.g., text and visuals that do not address items on pre- and posttest content knowledge measures vs. text and

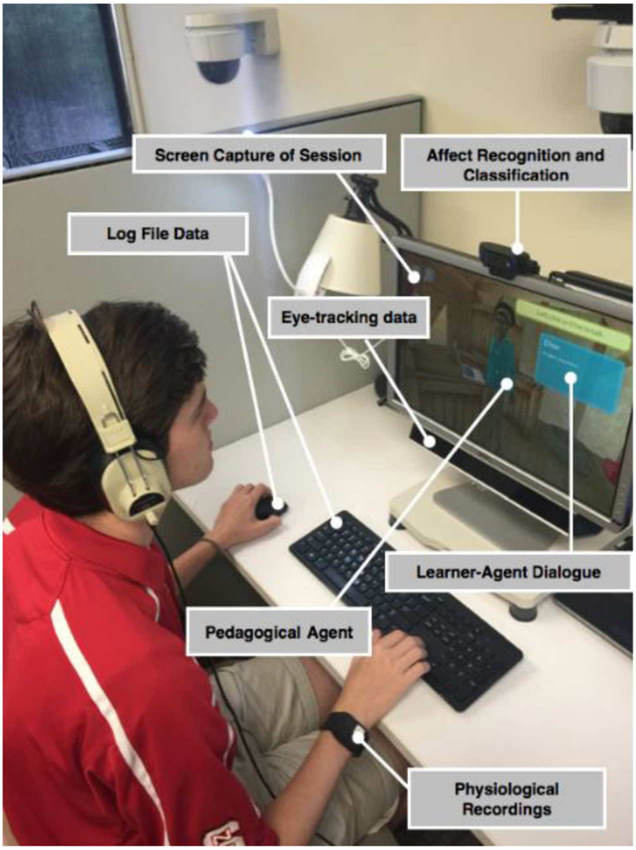


Fig. 4 Experimental set-up illustrating instrumented participant

visuals that directly address items on pre- and posttest content knowledge measures). NPCs also convey information that discuss domain knowledge. We characterized NPCs as interactions that contain both visuals and text. However, the visuals are the depiction of the NPC themselves where there is no information that is conveyed by the visual alone. The text serves as a dialogue between the NPC and the participant where the participant selects a predefined prompt pertaining to domain-specific content (e.g., "What is the smallest type of living organism?") and the NPC will respond via text and audio information related to the prompt. We operationalized books and research articles as text with no supporting visuals. Posters varied by how informative the information was to the participant. For example, the visuals provided in one poster may have conveyed information useful for acquiring domain knowledge (e.g. visual and supporting text of a cell wall) while another contained a visual with no related information to the domain (e.g., picture of a rainbow). We classified three types of presentations used during learning with Crystal Island: (1) NPCs (i.e., informative text with uninformative visuals), (2) books and research articles (i.e., informative text with no visuals), and (3) posters (i.e., informative text with a combination of informative and uninformative visuals).

Pretest Relevancy of Items

The individual items (e.g., specific book or NPC) of the different types of informational text presentations were separated into categories based on their relevance to items on the pretest. For example, the pretest question “What is the smallest type of living organism” is addressed by an NPC named Robert who explains that, “Bacteria are the smallest type of living organism.” This directly addresses the question, and therefore, interactions with Robert were labeled as relevant to the pretest. Between a total of 40 informational presentations, 47.5% of all types of presentations were relevant to pretest items (see Table 1). This categorization stems from priming literature (e.g., Mcnamara 2005) and assumes that participants should be more likely to identify information that relate to pretest answers as instructionally relevant as the learners have been pre-exposed to topics needed to successfully complete the posttest.

Normalized Change Scores

Learning gains from pretest to posttest were calculated using normalized change scores based on each participant’s pretest and posttest scores to mitigate learners’ pretest score biases (Marx and Cummings 2007). The normalized change scores captured changes in domain knowledge during the learning session with Crystal Island while controlling for the level of prior domain knowledge to ensure the scores were not biased towards participants who had greater prior knowledge about microbiology. Normalized change scores were calculated using Eqs. 1–3 depending on the difference between the pre- and posttest scores:

$$Normalized\ Change = \frac{post - pre}{100 - pre} \tag{1}$$

$$Normalized\ Change = 0 \tag{2}$$

$$Normalized\ Change = \frac{post - pre}{pre} \tag{3}$$

Table 1 Pretest relevancy item per type of presentation

Type of presentation	Total	Pretest-relevant	% of Pretest-relevant
NPCs	9	3	33.3%
Books & Research articles	21	12	57.1%
Posters	10	4	40.0%
All	40	19	47.5%

Equation 1 was applied when posttest scores were greater than pretest scores (i.e., the participant had more correct answers after the learning session and this included 70 [84%] participants in our sample). Equation 2 was applied when participants received the same score on the pretest and posttest (6 or 7% of the sample). Equation 3 was applied when participants had lower posttest scores relative to their pretest scores (7 participants or 8% of the sample). One participant with a normalized change score of 1 was removed from analyses as they were the only participant to receive a 100% on the posttest and which may have been caused by the ceiling effect (see Marx and Cummings 2007). The normalized change score was used in order to contextualize learning gains.

Eye-Tracking Data

Eye-tracking was used to identify fixation durations on AOIs (e.g., time spent fixating on text in books and research articles). In this study, we created predefined AOIs around the posters, books, research articles, and NPC visual/dialogue combined to identify the length of time participants fixated on the specific informational text presentations. A fixation in this study was operationally defined as a relatively still gaze on an AOI for a minimum of 250 ms (Salvucci and Goldberg 2000; see Fig. 5). The fixation durations in this study used seconds as a unit of measurement. Figure 5 provides visualizations of AOIs and fixations. The AOIs are represented by the colored shading over the NPC, dialogue box, book, and poster. The numbered circles show the order in which the fixations on AOIs occurred as well as a relative fixation duration indicated by the size of the indicator where a greater fixation duration is depicted by a larger circle.

Log-File Data

We used log-file data sequence and duration of participant interactions with game elements during learning with Crystal Island. Through these data, we identified different paths learners took and how they interacted with different game elements (i.e., books vs. NPCs). Specifically, when participants selected a poster, book, research article, or NPC, we captured and analyzed the frequency and time (in seconds) spent interacting with each of these game elements. Within this study, references made to log-file and eye-tracking data pertain to durations and fixation durations respectively. The differentiation between these two types of data are important in the connotation of cognitive processes that are represented by each. Log-file data indicate the opening and closing of an informational source (e.g., book). This denotes the selection process and the time spent within this source. Alternatively, eye-tracking data identifies fixations on

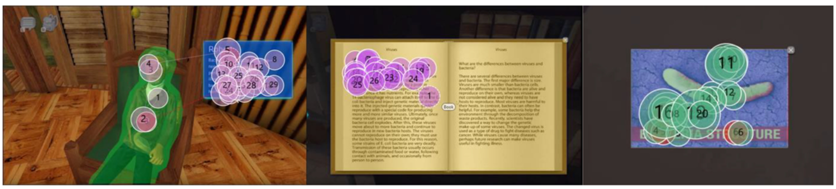


Fig. 5 Example visualizations for NPCs, books, and posters

the content of the source where this accounts for the participant looking off-screen or fixating on objects other than the content. As such, eye tracking denotes reading interactions whereas log files indicate overall time generally interacting with the object.

Outlier Removal

Outliers in the eye-tracking and log-file data were identified using boxplots and subsequently removed from the dataset. A total of 10 observations from the dataset containing 4779 observations from all 83 participants were removed, resulting in a total of 4769 observations between all participants.

Statistical Pre-Processing and Analysis

Data processing and analyses were completed using statistical programs Python (Python Core Team 2015) and R (RStudio Team 2018). Process data for analyses were collected and cleaned using Python. R was used to conduct statistical analyses. A three-level model used for this study was run using the 'lmer' and further analyzed with the 'analyze' functions using the 'lme4' (Bates et al. 2015) and 'psycho' (Makowski 2018) packages in R respectively.

Model Selection

The fitted three-level model uses three levels to explain differences in the data (see Fig. 6). This study utilized two models containing the same category of levels for each type of process data (e.g., eye-tracking, log-file). The third level for both models contains the participants who are separated based on their condition assignment. The second level contains the relevancy of level 1, types of informational text presentations, to answering the pretest items. The first level contains either the duration (captured from log files) or fixation duration (captured from eye-tracking) of the text, separating the two three-level models in terms of which process data each address. We identify the first level as being nested within pretest relevancy and the second level as being nested within participants. Maximum likelihood was used to fit both models to the data. From 83 participants, a total of 4769 observations were used to fit both models.

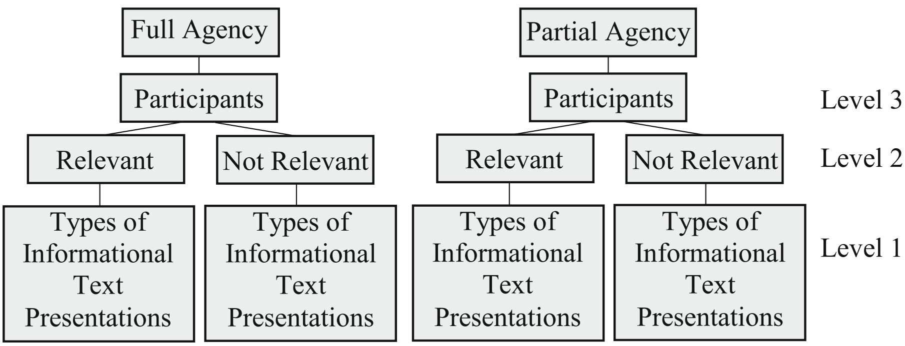


Fig. 6 Selected three-level model

Results

Preliminary Analyses

Preliminary analyses examined whether there were differences between informational text presentation process data to justify the analyses and interpretation of multiple types of data. Three paired t-tests were run to examine whether there were differences in the log-file and eye-tracking data between the types of presentations (see Table 2 for descriptive statistics). There was a significant difference, $t(1,82) = 6.73$, $p < .0001$, in the participant interactions with NPCs, where the durations of these instances were, on average, less than the fixation durations. Another paired t-test found a significant difference ($t(1,82) = 2.87$, $p < .01$) in participant interaction with books and research articles where participants had greater durations than fixation durations on these types of informational presentations. The third paired t-test identified differences in the types of data on posters ($t(1,82) = -12.0$, $p < .0001$) where, on average, durations were smaller than fixation durations.

In sum, these preliminary findings suggest that there are significant differences in log-file and eye-tracking data regarding informational presentations in the Crystal Island environment. From these findings, we may not conclude that the cognitive and metacognitive processes measured from log-file and eye-tracking data can be interchangeably modeled. For example, because we see that fixation duration on books and research articles were shorter than the log file durations of when these were open on the screen, it might suggest participants were fixating elsewhere on the screen or distracted from reading. The fixation durations for NPCs and posters were longer than the log file durations. This suggests that participants fixated on the NPCs and posters before selecting the NPC to talk to or the poster to read. As such, this suggests that log files may be a more accurate measure of participant interaction with information whereas eye tracking may be a more accurate measure of intention. Therefore, we must keep the log-file and eye-tracking data distinct in the further analyses to correctly understand and interpret how each channel of data contributes to participants' internal processes, external constraints, and subsequent interactions with the GBLE.

To address the presence of covariates within this study's analyses, learners' proportion of time spent on collecting information from NPCs, books and research articles, and posters as well as the frequency of these interactions are compared between conditions. First, two t-tests for log-file and eye-tracking data were run to examine how conditions differed in their proportion of time spent collecting all information

Table 2 Descriptive statistics of preliminary analyses of log-files and eye-tracking data

Type of informational text presentation	Log file		Eye tracking	
	M(sec)	*SD*(sec)	*M*(sec)	*SD*(sec)
NPCs	529.4	103.0	625.0	155.6
Books & Research articles	1829.5	554.3	1680.5	825.8
Posters	80.4	34.8	121.3	53.8

NPC Non-player characters

during their time within the environment. There were no differences between the full ($M = 0.46$, $SD = 0.07$) and partial ($M = 0.50$, $SD = 0.07$) agency conditions for duration proportion, $t(85.3) = 0.94$, $p > .05$. Similarly, full ($M = 0.46$, $SD = 0.14$) and partial ($M = 0.49$, $SD = 0.16$) agency conditions did not differ in their proportion of fixation durations, $t(96.8) = 0.90$, $p > .05$. Overall, the proportion of time collecting information does not vary between conditions.

Further analyses examined how conditions differed in their interactions with informational text presentations. Participants' interactions with informational text presentations were determined by the amount of control they were allowed within Crystal Island where participants in the partial agency were required to interact with all informational text presentations and those with full agency were not limited in their interactions. Three chi-squared tests were conducted to identify if participants differed in their informational text presentation frequency. Conditions did not vary in their frequency of NPC ($\mathcal{X}^2 = 19.3$, $p > .05$), book and research article ($\mathcal{X}^2 = 36.7$, $p > .05$), and poster ($\mathcal{X}^2 = 17.5$, $p > .05$) interactions (see Table 3). Overall, condition did not influence the frequency with which learners interacted with informational text presentations.

Research Question 1: Do Prior Knowledge and Learning Gains Significantly Differ Between Learners with Varying Levels of Autonomy?

For this research question, we included the no agency condition ($N = 32$) in addition to the full and partial agency conditions to analyze how participants with no autonomy learn with Crystal Island in comparison to participants afforded autonomy. We ran two ANOVAs for differences in both prior knowledge and normalized change scores between the three conditions. Results indicate that participants between each condition did not differ in their prior knowledge, $F(2,112) = 2.37$, $p > .05$. Further results indicated that those in the partial agency condition ($M = 0.45$, $SD = 0.27$) had significantly higher normalized change scores than those in the full agency condition ($M = 0.32$, $SD = 0.26$; $t(1,71.1) = 2.20$, $p = 0.03$) and participants in the no agency condition ($M = 0.12$, $SD = .26$; $t(1,64.8) = 5.14$, $p < .0001$; see Fig. 7). Participants with full agency had significantly greater normalized change scores than those in the no agency condition, $t(1,65.6) = 3.46$, $p < .001$. Overall, participants did not differ in prior knowledge between groups but did show a difference in their normalized change scores, with the partial agency condition learning more than learners with full and no autonomy. This suggests participants who were given restricted control (but therefore more scaffolding) over their choice of interacting with NPCs, read books and research articles, and

Table 3 Frequency of informational text presentations between conditions

Type of informational text presentation	Full agency		Partial agency	
	M	*SD*	*M*	*SD*
NPCs	18.7	5.98	19.9	5.23
Books & Research articles	22.2	6.77	27.3	7.66
Posters	13.2	4.51	15.1	4.29

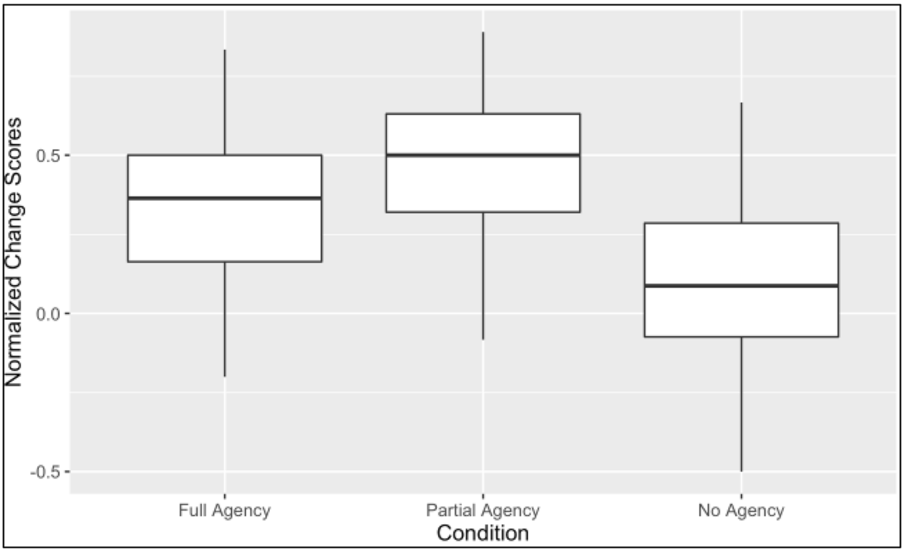


Fig. 7 Visualization of normalized change score differences between conditions

consult posters in the environment had a significantly greater learning outcomes than those given complete control or no control over their actions in the environment.

Research Question 2: Do Learners' Process Data for Each Type of Informational Text Presentation Predict Learning Gains?

Log-File Data

We first ran Pearson correlations for the participants' total content durations on each of the text presentations and normalized change scores (see Table 3). While total durations on NPCs and posters were not correlated with normalized change scores ($p > .05$), the total duration on books and research articles are significantly correlated with participants' normalized change scores, $r(83) = 0.27$, $p = 0.02$. In other words, the time a participant spent on books and research articles was significantly related to their learning.

A simple linear regression was calculated to identify if the total duration on books and research articles could predict participants' normalized change scores while controlling for participants' individual book and research article frequency and their proportion of time interacting with each type of informational text presentation. While durations on books and research articles are correlated with learning gains, when controlling for participants' frequency of book and research article interactions, these durations are unable to predict participants' normalized change scores.

Eye-Tracking Data

We calculated Pearson correlations for the participants' total content fixation durations on each of the text presentations and normalized change scores. Similar to the log-file data correlations, fixation durations on books and research articles was significantly correlated with participants' normalized change scores, $r(83) = 0.24$, $p = 0.03$, while fixation durations on NPCs and posters were not, $p > .05$. We calculated a simple linear regression to identify if total fixation duration on books and research articles could predict participants' normalized change scores while controlling for participants' book and research article frequency and their proportion of time interacting with each type of informational text presentation. While fixation durations on books and research articles are correlated with participants' normalized change scores, when controlling for participants' frequency of book and research article interactions, participants' fixation durations on these types of text presentations are unable to predict their learning gains ($p > 0.05$).

In sum, log-file (durations) and eye-tracking data (fixation durations) on books and research articles are significantly correlated with participants' normalized change scores in participants' while NPCs or posters are not. However, process data are not predictive of participants' normalized change scores. From these analyses, we question if participants' interactions with text presentations measured by process data is related to the quality of the information within the NPCs, books and research articles, and posters and the autonomy participants are afforded within the environment.

Research Question 3: Do Learners' Varying Levels of Autonomy Influence How Learners Interact with Each Type of Informational Text Presentation?

Two separate two-way repeated measures ANOVAs for log-file and eye-tracking data were run to analyze if there were significant differences between full and partial agency conditions and within the different types of informational text presentations. For these analyses, condition containing two levels (i.e., full agency, partial agency) represent the between-subjects effect while the type of informational text presentation is the between-subjects factor with each level (i.e., NPCs, books and research articles, posters) are repeatedly measured for each participant as they complete the game.

Log-File Data

A two-way repeated measures ANOVA was run to identify the differences between condition within the different types of informational text presentations using log-file durations as an outcome variable. Results indicated a significant main effect of condition, $F(1,81) = 21.2$, $p < .0001$, and type of informational text presentation on durations, $F(2,162) = 813.0$, $p < .0001$. Results for the content durations yielded a significant interaction value, $F(2,162) = 11.5$, $p < .0001$, between the condition and the type of information presentation. Post-hoc pairwise comparisons of the within-subjects factor supported significant differences between the content durations (i.e., types of text presentation; $p < .0001$). Specifically, participants interacted with books and research articles ($M = 1829.5$ s, $SD = 554.3$ s) for a longer duration than NPCs ($M = 529.4$, $SD = 103.0$) and posters ($M = 80.4$ s, $SD = 34.8$ s). All post- hoc pairwise comparisons of the interaction effect (condition and type of text presentation) indicated significant results using a Bonferroni correction ($p < .017$; see Table 4) where participants in the partial agency had greater durations than those in the full agency condition.

Table 4 Descriptive statistics and post-hoc comparisons of each process data

Process data	Type of presentation	Full agency		Partial agency		
		M(sec)	*SD*(sec)	*M*(sec)	*SD*(sec)	*t*
Log-file	NPCs	491.6	101.1	581.1	81.7	−46.8*
	Books & Research Articles	1643.3	526.8	2084.0	491.4	−30.1*
	Posters	63.9	29.9	103.2	27.6	−20.9*
Eye-tracking	NPCs	584.3	155.0	680.8	140.2	−36.6*
	Books & Research Articles	1513.8	727.6	1909.1	905.4	−18.5*
	Posters	105.3	55.7	143.4	42.7	−20.5*

*$p < .0001$

Eye-Tracking Data

A two-way repeated measures ANOVA was run to identify the differences between condition within the different types of informational text presentations using eye-tracking fixation durations as an outcome variable. Findings indicated a significant relationship between fixation duration and condition, $F(1,81) = 6.92$, $p < .05$, as well as type of informational text presentation, $F(2,162) = 261.2$, $p < .0001$. Results for content fixation durations yielded a significant interaction value, $F(2,162) = 3.70$, $p < .05$, between the condition and the type of information presentation. Pairwise comparisons supported significant differences between the content durations of the within-subjects factor (i.e., types of text presentation; $p < .0001$) where fixations on books and research articles ($M = 1680.5$ s, $SD = 825.8$ s) are greater than those on NPCs ($M = 625.0$ s, $SD = 155.6$ s) and posters ($M = 121.3$ s, $SD = 53.8$ s). Post-hoc analyses with a Bonferroni correction ($p < .05$) identified significant pairwise comparisons between the levels of each factor (see Table 4) where those in the partial agency tended to have greater fixations on each type of presentation than those in the full agency.

In sum, log-file and eye-tracking data report similar findings where there are differences in the durations and fixation durations between conditions themselves, indicating that the total duration and fixation duration spent interacting with informational material in the environment differs between condition where participants in the partial agency had consistently greater durations and fixation durations than those in the full agency (see Table 4). When we introduce the type of presentation as a within-subjects factor, we can identify differences in the time spent and fixated on different types of presentations between the two conditions.

Research Question 4: Do Learners' Varying Levels of Autonomy and the Relevancy of Informational Text Influence How Learners Interact with Each Type of Presentation?

Conditional three-level growth models were run for process data (Fig. 8).

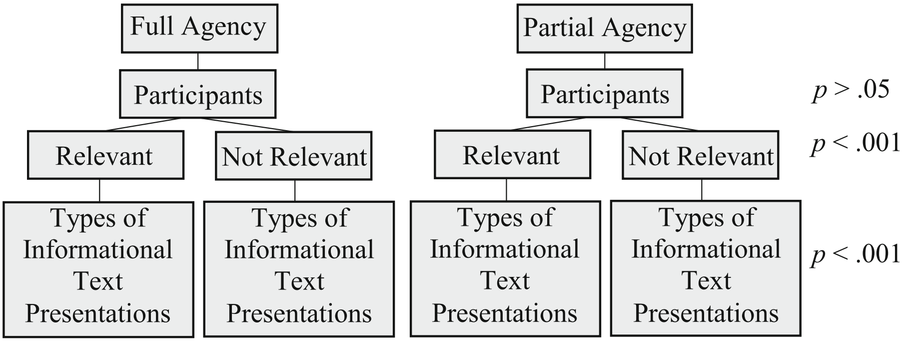


Fig. 8 Three-level model used for process data

Log-File Data

The first model examining differences in duration failed to converge, meaning that the model that was run does not fit the data well. Therefore, we disregarded this model for further analyses in order to accurately interpret the effects of condition, support of the pretest, and type of informational text presentations.

Eye-Tracking Data

The second model examining differences in fixation durations has a total explanatory of 40.16% where we may then examine the effects of the type of informational text presentations, relevance of the pretest questions, condition, and their interaction with each other. The relevancy of pretest questions significantly contributes to this model, $t(4686) = 11.3$, $p < .001$; std. $\beta = 0.62$. There is a significant difference between groups of pretest relevancy where participants fixated significantly more on pretest-relevant texts ($M = 50.5$ s, $SD = 58.3$ s) than pretest-irrelevant texts ($M = 35.1$ s, $SD = 48.2$ s), not accounting for the type of informational text presentation. In addition to this main effect and in support of previous analyses, books and research articles, $t(123) = 12.2$, $p < .001$; std. $\beta = 1.03$, and posters, $t(1377) = -4.36$, $p < .001$; std. $\beta = -0.23$, significantly contribute to the prediction of fixation durations (using NPCs as a reference variable), where fixation durations on books and research articles ($M = 69.0$ s, $SD = 65.3$ s), NPCs ($M = 32.6$ s, $SD = 37.1$ s), and posters ($M = 8.7$ s, $SD = 5.45$ s) significantly differ from each other. There is an interaction effect between types of informational text presentations and pretest relevancy (Table 5) where posters and pretest relevancy, $t(4687) = -7.36$, $p < .001$; std. $\beta = -0.62$, as well as books and research articles and pretest relevancy, $t(4615) = -10.5$, $p < .001$; std. $\beta = -0.77$, using NPC as a reference variable, significantly contribute to the model's overall explanatory power of fixation durations. Interestingly, condition does not significantly contribute to the overall model predicting fixation durations when separating informational text presentations based on their contribution and relevance to the pretest.

In sum, differences in durations are not able to be accurately identified through a complex model which takes into account repeated measures nested within multiple levels. Differences in fixation durations, however, are able to be identified within types of informational text presentations as well as whether or not the type of presentation

Table 5 Descriptive statistics of pretest-relevant groups and types of informational text presentation fixation durations

Pretest relevancy group	Type of presentation	*N*	*M*(sec)	*SD*(sec)
Relevant	NPC	579	54.3	50.2
	Books & Research Articles	1150	66.3	65.3
	Posters	487	8.53	5.76
Not relevant	NPCs	1101	20.2	17.4
	Books & Research Articles	871	72.6	65.2
	Posters	671	8.82	5.21

item was relevant to information in the pretest that participants were exposed to prior to interacting with Crystal Island. Participants, regardless of autonomy afforded, interacted with more pretest-relevant text. However, looking at the descriptive statistics (see Table 5), participants fixated longer on NPCs that were relevant to the pretest, but fixated for a greater period of time on books and research articles as well as posters which were not relevant to pretest answers. This could be used as a proxy addressing learners' competencies in metacognitive judgments where content evaluations were accurate relating to information provided by NPCs, but not information presented by the other types of presentations.

Discussion

In this study, we investigated types of informational text presentations and autonomy to assess the impact of learning and metacognitive process use within GBLEs captured through multiple types of data (i.e., eye-tracking, log-files). Preliminary results indicated that log-file and eye-tracking data significantly differed from each other when considering the overall time spent on different types of informational text presentations. Generally, eye-tracking data (i.e., fixation durations on AOIs) was found to have longer durations than those reported in log-files (i.e., durations of presentation instances) with the exception of book and research article interactions. Although seemingly contradictory, this is a result of the environment itself where learners could look at posters and NPCs before interacting with them as indicated by log files. This addresses the need for researchers to consider multiple types of data streams and analyses for different GBLE features. These preliminary analyses further suggest a need for future studies to examine which types of data accurately capture learner element interactions within GBLEs and learners' cognitive and metacognitive skills that are demonstrated through these interactions. However, based on the findings of the three-level model fit to durations and fixation durations, we suspect that fixation duration on posters, books, and research articles captured cognitive processes more accurately compared to log-file data, which captured cognitive processes during NPC interactions more accurately. Additional preliminary analyses confirmed that conditions did not differ in the proportion of time and the frequency of instances for each type of informational presentation. As a result, when conditions were compared, these factors were not included as covariates.

Overall Findings

Results from our *first* research question confirmed our hypothesis. There were no differences in the level of prior knowledge between all three agency conditions but suggested significant differences in learning gains between the conditions, where those with restricted control (i.e., in the partial agency condition) over their actions had higher learning gains compared to learners with full control. Learners with no agency had significantly lower learning gains than those in the full and partial agency conditions. This is partially consistent with previous research studies showing that limited autonomy within GBLEs increases learning outcomes (Bradbury et al. 2017; Sabourin et al. 2013). However, this emphasizes the need to moderate the autonomy afforded to learners where learners who are too restricted (e.g., no agency) and learners who have full control (e.g., full agency) exemplify that extreme levels of autonomy do not outperform learners with a moderate amount of control (e.g., partial agency). More specifically, this finding identifies autonomy, moreover restricted autonomy, as a scaffold which supports learning through directing learner interactions with informational content. As the availability of the text presentations within Crystal Island was consistent regardless of conditions, where learners were able to interact with the same amount of texts that share the same content, we must question why participants with total control over their actions did not engage with all information available to them through the environment.

Next, we investigated whether there were differences in the interactions of types of text presentations and how these presentations relate to overall learning between agency conditions. Results partially confirmed our hypotheses where time spent fixating and opening books and research articles were correlated with higher normalized change scores, while NPCs and posters do not. We initially expected that duration and fixation duration on types of text that contain both diagram and text would positively predict learning gains. The finding that fixation and fixation duration on NPCs and posters did not positively predict learning gains is misaligned with the CTML framework (Mayer 2014). According to this framework, diagrams and text presented simultaneously, represented in this paper as NPCs and posters, were predicted to result in greater learning rather than books and research articles. Our results showed that even though durations and fixation durations on types of informational text presentations while controlling for proportion of time and frequency of interactions cannot predict learning gains, these process data for books and research articles are significantly correlated with learning gains. We posit that participants valued the context-rich information more than the presentation of a non-informative diagram alongside informative information. Future studies should examine if informative diagrams (e.g., physical symptoms of the illness) alongside relevant informative information (i.e., dialogue) are predictive of learning gains.

Our *third* research question investigated the differences in the duration and fixation duration between condition within types of informational text presentations. Findings confirmed our hypotheses and results suggested there were differences in total time spent on all text presentations and fixating on informational content between agency conditions, where learners with partial control over their actions (but consequently more scaffolding) consistently have greater durations and fixation durations on all types of presentations than those with full control. From the results, we see a difference in the overall time spent reading the types of informational text presentations where participants spent more time on books and research articles than any other presentation of information, followed by NPCs and then posters. From this we conclude that participants valued the context-rich information provided by books and research articles, closely following the conclusions from the previous research question. Examining the interaction between condition and types of informational text presentations utilizing process data, we see that there is a significant interaction where, accounting for each type of presentation, there were differences between the two conditions where partial agency had consistently greater durations and fixation durations than full agency. It is important to note that these findings are consistent with preliminary findings where, with the exception of NPCs, fixation durations on the presentation is less than durations.

Finally, we investigated the interaction of durations and fixation durations between condition, pretest-relevant items, and types of informational text presentations. This research question aimed to identify how learners used metacognitive processes within GBLEs which either limit or allow learners' control over their interactions with the environment. Overall results partially confirmed our hypotheses where we initially expected to see a difference between conditions as previous as previous studies have shown that autonomy is a detriment to learners' metacognitive judgements (e.g., content evaluations; Bradbury et al. 2017; Azevedo et al. 2019). However, the first model looking at durations would not converge, indicating that an accurate model could not be identified with log-file data. Our fitted model, using eye-tracking data, nested the items and presentations within participants. Results showed significant differences in the time spent fixating on books and research articles, NPCs, and posters, differences in fixation durations related to the relevancy of pretest items, with a significant interaction between these two factors, but there were no differences between agency conditions. Regardless of condition, there are differences in selecting and utilizing context in text presentations related to pretest items. Overall, participants fixated on content in text presentations related to pretest-relevant items more than irrelevant types of informational text presentations. This demonstrates participants employing metacognitive strategies as approximately half of the total number of informational text within the environment were relevant to the pretest and therefore, participants would not display a difference in their fixation durations if metacognitive strategies were not used by participant. However, when accounting for the type of presentation, NPCs are fixated on more for pretest-relevant than pretest-irrelevant information whereas greater fixations are spent on pretest-irrelevant books and research articles and posters. This finding indicates that learners, regardless of autonomy, make accurate metacognitive judgments when encountering NPCs than any other type of informational text presentation. This may be due to the content of NPCs which contain a diagram and information that is presented in a more conversational method. From these overall findings, we conclude that for all types of informational text presentations, there is a need for an increase in scaffolding to fully support learners in their metacognitive judgments.

Application to CTML

Our findings highlight the need for CTML to be integrated into GBLEs that provide multiple presentations of informational content critical for learning. Specifically, our results show that learning is impacted by how information is represented. In Mayer's (2014) CTML framework, information related to the overall goal of learning is provided to learners, where learners are then required to evaluate individual sources of information. Alternatively, within GBLEs, learners must first seek out sources of information throughout their interaction with the game elements before identifying how relevant information within individual sources are related and relevant within each other (i.e., visual and text) and between different sources in relation to game completion and learning.

The current study's results support modifying the CTML model to consider how various features, elements, and goals of narrative-centered GBLEs impact learners' metacognitive judgments in selecting relevant information and their impact on learning gains. Features generalizable to all GBLEs include the autonomy afforded to learning which may be manipulated within the environment, ultimately changing the control learners have over their learning with the intention of optimizing their learning. Elements, including the type of text presentations, directly address CTML in how learners must select, organize, and integrate multiple sources of information that are presented in several ways using text and diagrams.

Within this study, limited autonomy prompted the participant to interact with all types of informational text presentations, forcing the exposure of all content related to domain knowledge including the specific knowledge items represented on the pretest. Given the results of the current study referencing the autonomy feature of narrative-centered GBLEs, we conclude that autonomy captures context-relevant sources of information and this type of scaffold influences domain content knowledge where a greater amount of autonomy ultimately resulted in lower learning gains than restricted autonomy. Modifications to CTML are proposed to fully apply this cognitive theory to GBLEs that contain informational content. We propose modifying CTML using the three condensed phases of cognitive processes: 1) selecting; 2) organizing; and 3) integrating content from an informational source. We identify two levels within the *selecting* phase of CTML which may vary dependent upon the level of autonomy allowed to learners. There are two proposed levels within the selection phase: 1) GBLE Information Presentation; and 2) Content of Information Presentation. The first level refers to the unique need in GBLEs to identify and select information presentations (e.g., scientific books and research articles) which relate to domain-specific content, or content related to the pretest, but also to the overarching goal of the game itself where learners must identify information relevancy in reference to the domain content they must know or the information needed to complete the immediate goal of the GBLE. However, this level may include elements that do not contribute to the knowledge needed to complete the posttest or the GBLE itself. For example, the way in which learners synthesize information with the Crystal Island worksheet could influence how learners identify and select relevant information. The second level refers to the traditional use of CTML where, given a large chunk of information, learners must select relevant information which will increase learning. This emphasizes the need for the learner to identify relevant information across multiple documents and types of informational text presentations. This modification to Mayer's CTML would allow for learners' metacognitive judgments of informational text within GBLEs to be evaluated at multiple levels.

Limitations and Future Directions of the Study

We acknowledge limitations in our study related to the interpretation of analyses, classification of groups, and potentially influential factors. It is first important to note that in the simple linear regression models (Research Question 2), given our positive intercept beta, our model fails to correctly capture when a participant performs worse on posttest compared to the pretest performance. This is due to the fact that our predictor variable must be positive as a participant cannot look at informational text presentations for a negative amount of time. Although this is a limitation to the interpretation of analyses, it raises the question of how a learner would perform worse on the posttest. Specifically, did the participant guess on answers on the pretest and were incorrect for the congruent form of the question on the posttest? If so, this would be an example of a participant not learning compared to the "unlearning" if analyses were to be interpreted using negative values. Alternatively, did the environment prove to be too distracting and only serve to confuse certain students resulting in worse performance on the posttest?

Limitations of this study include the identification of item pretest relevancy. The relevancy of NPCs was determined by a singular piece of information shared by the NPC, where other information shared may not have been relevant. For example, learners have the choice of selecting 3 questions for the NPC to answer. If one of the three questions contained pretest-relevant information, regardless of the relevancy of the other questions, that NPC was determined to be relevant to the pretest items rather than the question itself.

Further limitations include the exclusion of additional influences of time spent in the environment itself. Although the time in the environment varied between conditions, the time exposed to the different types of informational text presentations in comparison to overall time in game did not differ between conditions. As such, these analyses ignore the duration of time participants spent on concept matrices, worksheet edits, and scanning food items. However, as the goal of this current study was to identify how autonomy and metacognitive judgements influenced learners' interactions with informational text in GBLEs measured by process data, the inclusion of these elements was outside the scope of the study. Further, the three game elements investigated within this study were the only sources of information that could contribute to learners' domain knowledge for items tested in the posttest and, therefore, are the instrumental elements in examining learning gains that should be investigated within this study.

Additionally, this study did not account for how much information is included in each presentation where books and research articles contained significantly more information than posters. However, these times are compared with each other as the same metacognitive processes are utilized to read, select, organize, and integrate information within each source. Although this may be considered a limitation of the study, this may be a limitation of the environment itself as these types of presentations which do not offer rich text and do not influence learning gains may be considered distractions from the presentations which promote learners' domain knowledge. Despite these potential limitations, this study incorporates all elements that are critical to identifying how learners interact with informational text presentations within GBLEs.

While limitations are identified, we acknowledge multiple directions in which this study can expand. This study emphasizes the need for metacognitive processes to be integrated within cognitive literature, and as such, it is important to mention conditions driving individual learners' interaction with GBLEs and implementation of these processes. For example, learners' executive functions, cognitive resources, motivation, emotions, prior knowledge, etc. can influence how learners interpret the task, or environment, and implement metacognitive and cognitive processes while learning with a GBLE (McCardle and Hadwin 2015; Winne 2018). Future directions of investigation may also include comparing an open-ended learning environment, such as a GBLE, to simulations which are a more restricted environment (e.g., O'Keefe et al. 2014). In examining these comparable environments, researchers can understand the effect of a more restricted environment with a limited number of resources from which the learners receive information.

Future Directions in AI and GBLEs

This study supports the need for artificial intelligence interfaces to include adaptable scaffolds using autonomy which functions off of eye-tracking and log-file data which changes depending upon the type of informational text that is being presented. Generally, this applies to GBLEs which focus on presenting complex instructional multimedia content related to topics and STEM domains. Given the results of this study, we identify the need for GBLEs to intelligently and adaptively support learners throughout the environment as they use metacognitive judgements to select information deemed relevant to learning outcomes. The aim of integrating AI with GBLEs stems from the need for scaffolding where learners are given support through restricted autonomy to address the challenge of accurately applying metacognitive processes. Indicated by results from this study, learners are not able to accurately identify and select relevant information. Therefore, AI within GBLEs should address the need for a constantly adaptable scaffold for autonomy to fade in and out throughout learner interactions with the environment dependent upon the individualized (and contextualized) process data of each learner. Additionally, this should account for the type of informational text presentation as, according to the preliminary analyses, different types of process data should be used for each type of learning resource and depending on how they are utilized by learners. Future iterations of Crystal Island, as well as other GBLEs, should include greater AI to (a) support metacognitive judgements by assisting learners in selecting, organizing, and integrating informational sources and content within these sources and (b) integrate adaptive scaffolding dependent upon the real-time feedback from individual process data.

Acknowledgements This research was supported by funding from the Social Sciences and Humanities Research Council of Canada (SSHRC 895-2011-1006). Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Social Sciences and Humanities Research Council of Canada. The authors would also like to thank members of the SMART Lab and the intelliMEDIA group at NCSU for their assistance and contributions.

References

Akram, B., Min, W., Wiebe, E., Mott, B., Boyer, K. E., & Lester, J. (2018). Improving stealth assessment in game-based learning with LSTM-based analytics. In K. E. Boyer & M. Yudelson (Eds.), *EDM'18: Proceedings of the 11th international conference on Educational Data Mining* (pp. 208–218). Buffalo: EDM.

Alghamdi, E. A. (2016). Untangling multimedia effects on EFL incidental vocabulary learning via playing an online hidden-object game. *International Journal of Computer-Assisted Language Learning and Teaching, 6*, 24–40.

Alonso-Fernández, C., Cano, A. R., Calvo-Morata, A., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2019). Lessons learned applying learning analytics to assess serious games. *Computers in Human Behavior, 99*, 301–309.

Azevedo, R., & Gasevic, D. (2019). Analyzing multimodal multichannel data about self-regulated learning with advanced learning technologies: issues and challenges. *Computers in Human Behavior, 96*, 207–210.

Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition – implications for the design of computer-based scaffolds. *Instructional Science, 33*, 365–379.

Azevedo, R., & Taub, M. (2020). The challenge of measuring processes and outcomes while learning from multiple representations with advanced learning technologies. In P. Van Meter, A. List, D. Lombardi, & P. Kendeou (Eds.), *Handbook of learning from multiple representations and perspectives* (pp. 532–553). Cambridge: Cambridge University Press.

Azevedo, R., Taub, M., & Mudrick, N. V. (2018). Understanding and reasoning about real-time cognitive, affective, and metacognitive processes to foster self-regulation with advanced learning technologies. In D. H. Schunk & J. A. Greene (Eds.), *Handbook on self- regulation of learning and performance* (2nd ed., pp. 254–270). New York: Routledge.

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

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

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48.

Bradbury, A., Taub, M., & Azevedo, R. (2017). The effects of autonomy on emotions and learning in game-based learning environments. In R. Granger, U. Hahn, & R. Sutton (Eds.), *CogSci'17: Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 1666–1671). London: Cognitive Science Society.

Burkett, C. & Azevedo, R. (2012). The effect of multimedia discrepancies on metacognitive judgments. *Computers in Human Behavior, 28*, 1276–1285.

Butcher, K. (2014). The multimedia principle. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 174–205). Cambridge: Cambridge University Press.

Chen, Y., & Tsai, M. J. (2015). Eye-hand coordination strategies during active video game playing: an eye-tracking study. *Computers in Human Behavior, 51*, 8–14.

Chen, Z.-H., Chen, H. H.-J., & Dai, W.-J. (2018). Using narrative-based contextual games to enhance language learning: a case study. *Journal of Educational Technology & Society, 21*, 186–198.

Cheng, M. T., Lin, Y. W., & She, H. C. (2015). Learning through playing virtual age: exploring the interactions among student concept learning, gaming performance, in-game behaviors, and the use of in-game characters. *Computers & Education, 86*, 18–29.

Cierniak, G., Gerjets, P., & Scheiter, K. (2009). Expertise reversal in multimedia learning: Subjective load ratings and viewing behavior as cognitive process indicators. In N. A. Taatgen & H. van Rijn (Eds.), *CogSci'09: Proceedings of the 31st Annual Meeting of the Cognitive Science Society* (pp. 1906–1911). Austin: Cognitive Science Society.

D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: a gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies, 70*, 377–398.

Dever, D. A., & Azevedo, R. (2019a). Autonomy and types of informational text presentations in game-based learning environments. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *AIED'19: Proceedings of the 20th International Conference on Artificial Intelligence in Education* (pp. 110–120). Amsterdam: Springer.

Dever, D. A., & Azevedo, R. (2019b). Examining gaze behaviors and metacognitive judgments of informational text within game-based learning environments. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *AIED'19: Proceedings of the 20th International Conference on Artificial Intelligence in Education* (pp. 121–132). Amsterdam: Springer.

Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2019). Detecting mistakes in CPR training with multimodal data and neural networks. In R. G. Crespo & D. Burgos (Eds.), Sensors*,* 19, 3099–3119.

Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: critique, illustration, and applications. *Journal of Educational Psychology, 100*, 613–628.

Emerson, A., Sawyer, R., Azevedo, R., & Lester, J. (2018). Gaze-enhanced student modeling for game-based learning. In J. Zhang & T. Mitrovic (Eds.), *UMAP'18: Proceedings of the 26th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 63–72). New York: ACM.

Giannakos, M. N., Sharma, K., Pappas, I. O., Kostakos, V., & Velloso, E. (2019). Multimodal data as a means to understand the learning experience. *International Journal of Information Management, 48*, 108–119.

Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P., & Kreuz, R. (1999). AutoTutor: a simulation of a human tutor. *Journal of Cognitive Systems Research, 1*, 35–51.

Greene, J., & Azevedo, R. (2009). A micro-level analysis of SRL processes and their relations to the acquisition of a sophisticated mental model of a complex system. *Contemporary Educational Psychology, 34*, 18–29.

Greene, J., Bolick, C., & Robertson, J. (2010). Fostering historical knowledge and thinking skills using hypermedia learning environments: the role of self-regulated learning. *Computers & Education, 54*, 23–243.

Herrington, J., Reeves, T. C., & Oliver, R. (2014). Authentic learning environments. In M. D. Spector, J. Merrill, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (pp. 401–412). New York: Springer.

iMotions Attention Tool (Version 6.0) [Computer software] (2016). Boston: iMotions Inc.

Kiili, K., Ketamo, H., & Kickmeier-Rust, M. D. (2014). Evaluating the usefulness of eye tracking in game-based learning. *International Journal of Serious Games, 1*, 51–65.

Lai, M. L., Tsai, M. J., Yang, F. Y., Hsu, C. Y., Liu, T. C., Lee, S. W. Y., Lee, M. H., Chiou, G. L., Liang, J. C., & Tsai, C. C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review, 10*, 90–115.

Lee, S. Y., Mott, B. W., & Lester, J. (2011, June). Modeling narrative-centered tutorial decision making in guided discovery learning. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), *AIED'11: Proceedings of the 15th International Conference on Artificial Intelligence in Education* (pp. 163–170). New York: Springer.

Lin, S. S. (2014). Cognitive load for configuration comprehension in computer-supported geometry problem solving: an eye movement perspective. *International Journal of Science and Mathematics Education, 12*, 605–627.

Makowski, D. (2018). The psycho package: An efficient and publishing-oriented workflow for psychological science. In T. J. Leeper (Ed.), Journal of Open Source Software, 3, 470–471.

Marx, J. D., & Cummings, K. (2007). Normalized change. *American Journal of Physics, 75*, 87–91.

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). Cambridge: Cambridge University Press.

Mayer, R. E. (2019). Computer games in education. *Annual Review of Psychology, 70*, 531–549.

McCardle, L., & Hadwin, A. F. (2015). Using multiple, contextualized data sources to measure learners' perceptions of their self-regulated learning. *Metacognition and Learning., 10*, 43–75. https://doi.org/10.1007/s11409-014-9132-0.

Mcnamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. New York: Psychology Press.

O'Keefe, P. A., Letourneau, S. M., Homer, B. D., Schwartz, R. N., & Plass, J. L. (2014). Learning from multiple representations: an examination of fi xation patterns in a science simulation. *Computers in Human Behavior, 35*, 234–242.

Park, H., Lee, S., Lee, M., Chang, M. S., & Kwak, H. W. (2016). Using eye movement data to infer human behavioral intentions. *Computers in Human Behavior, 63*, 796–804.

Pea, R. R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences, 13*, 423–451.

Pekrun, R., Goetz, T., Frenzel, A., Barchfeld, P., & Perry, R. (2011). Measuring emotions in students' learning and performance: the achievement emotions questionnaire (AEQ). *Contemporary Educational Psychology, 36*, 36–48.

Plass, J. L., Homer, B. D., Kinzer, C. K., Chang, Y. K., Frye, J., Kaczetow, W., Isbister, K., & Perlin, K. (2013). Metrics in simulations and games for learning. In A. Drachen, M. S. El-Nasr, & A. Canossa (Eds.), *Game analytics* (pp. 697–729). New York: Springer.

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

Plass, J., Mayer, R. E., & Homer, B. (Eds.). (2020). *Handbook of game-based learning*. Cambridge: MIT Press.

Rowe, J. P., McQuiggan, S. W., Robison, J. L., & Lester, J. C. (2009). Off-task behavior in narrative-centered learning environments. In S. Craig & D. Dicheva (Eds.), *AIED'09: Proceedings from the 14th International Conference on Artificial Intelligence in Education* (pp. 99–106). Berlin, Germany: Springer-Verlag.

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

RStudio Team (2018). RStudio: Integrated development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

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

Sabourin, J., Shores, L., Mott, B., & Lester, J. (2013). Understanding and predicting student self-regulated learning strategies in game-based learning environments. *International Journal of Artificial Intelligence in Education, 23*, 94–114.

Salvucci, D., & Goldberg, J. (2000). Identifying fixation and saccades in eye-tracking protocols. In A. T. Duchowski (Ed.), *ETRA'00: Proceedings of the 2000 symposium on Eye Tracking Research & Applications* (pp. 71–78). New York: ACM.

Schraw, G., Bruning, R., & Svoboda, V. (1995). Sources of situational interest. *Journal of Literacy Research, 27*, 1–17.

Serge, S. R. (2014). A multimedia approach to game-based training: Exploring the effects of the modality and temporal contiguity principles on learning in a virtual environment (Doctoral dissertation). University of Central Florida, Orlando, USA.

Shih, J. L., Jheng, S. C., & Tseng, J. J. (2015). A simulated learning environment of history games for enhancing players' cultural awareness. *Interactive Learning Environments, 23*, 191–211.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. *Computer Games and Instruction, 55*, 503–524.

Spires, H. A., Rowe, J. P., Mott, B. W., & Lester, J. C. (2011). Problem solving and game-based learning: effects of middle grade students' hypothesis testing strategies on learning outcomes. *Journal of Educational Computing Research, 44*, 453–472.

Taub, M., Mudrick, N., Azevedo, R., Millar, G., Rowe, J., & Lester, J. (2016). Using multi-level modeling with eye-tracking data to predict metacognitive monitoring and self-regulated learning with Crystal Island. In A. Micarelli, J. Stamper, & K. Panourgia (Eds.), *ITS'16: Proceedings of the thirteenth International Conference on Intelligent Tutoring Systems* (pp. 240–246). New York: Springer.

Taub, M., Mudrick, N. V., Azevedo, R., Millar, G. C., Rowe, J., & Lester, J. (2017). Using multi-channel data with multi-level modeling to assess in-game performance during gameplay with Crystal Island. *Computers in Human Behavior, 76*, 641–655.

Taub, M., & Azevedo, R. (2018). Using sequence mining to analyze metacognitive monitoring and scientific inquiry based on levels of efficiency and emotional expressivity during game-based learning. *Journal of Educational Data Mining, 10*, 1–26.

Taub, M., Azevedo, R., Bradbury, A. E., Millar, G. C., & Lester, J. (2018). Using sequence mining to reveal the efficiency in scientific reasoning during STEM learning with a game-based learning environment. *Learning and Instruction, 54*, 93–103.

Taub, M., Azevedo, R., Rajendran, R., Cloude, E. B., Biswas, G., & Price, M. J. (in press). How are students' emotions related to the accuracy of cognitive and metacognitive processes during learning with an intelligent tutoring system? *Learning and Instruction.*, 101200. https://doi.org/10.1016/j.learninstruc.2019.04.001.

Team, P. C. (2015). *Python: A dynamic, open source programming language*. Python Software Foundation.

Tsai, M. J., Huang, L. J., Hou, H. T., Hsu, C. Y., & Chiou, G. L. (2016). Visual behavior, flow and achievement in game-based learning. *Computers & Education, 98*, 115–129.

Winne, P. H. (2018). Theorizing and researching levels of processing in self-regulated learning. *British Journal of Educational Psychology, 88*, 9–20.

Winne, P. H., & Hadwin, A. F. (2013). nStudy: Tracing and supporting self-regulated learning in the internet. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 293–308). New York: Springer.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: a presence questionnaire. *Presence: Teleoperators and Virtual Environments, 7*, 225–240.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Daryn A. Dever[1] **· Roger Azevedo**[1] **· Elizabeth B. Cloude**[1] **· Megan Wiedbusch**[1]

✉ Daryn A. Dever
ddever@knights.ucf.edu

Roger Azevedo
roger.azevedo@ucf.edu

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

Megan Wiedbusch
meganwiedbusch@knights.ucf.edu

[1] Department of Learning Sciences and Educational Research, University of Central Florida, Orlando, FL, USA