Discourse Processes

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hdsp20

Emotions and the Comprehension of Single versus Multiple Texts during Game-based Learning

Daryn A. Dever, Megan D. Wiedbusch, Elizabeth B. Cloude, James Lester & Roger Azevedo

DISCOURSE PROCESSES
2022, VOL. 59, NOS. 1–2, 94–115
https://doi.org/10.1080/0163853X.2021.1950450

Routledge
Taylor & Francis Group

Emotions and the Comprehension of Single versus Multiple Texts during Game-based Learning

Daryn A. Dever[a], Megan D. Wiedbusch[a], Elizabeth B. Cloude[b], James Lester[c], and Roger Azevedo[a]

[a]School of Modeling, Simulation, and Training, University of Central Florida, Florida; [b]Department of Learning Sciences and Educational Research, University of Central Florida, Florida; [c]Department of Computer Science, North Carolina State University, North Carolina

ABSTRACT

This study examined 57 learners' emotions (i.e., joy, anger, confusion, frustration) as they engaged with scientific content while learning about microbiology with Crystal Island, a game-based learning environment (GBLE). Measures of learners' prior knowledge, in-game text comprehension, facial expressions of emotion, and posttest reading comprehension were collected to examine the relationship between emotions and single- and multiple-text comprehension. Analyses found that both discrete and non-discrete emotions were expressed during reading and answering in-game assessments of single-text comprehension. Learners expressed greater joy during reading and greater expressions of anger, confusion, and frustration during in-game assessments. Further results found that learners who expressed a high number of different emotions throughout reading and completing in-game assessments tended to have lower in-game comprehension scores whereas a higher number of different expressed emotions while completing in-game assessments was associated with greater posttest comprehension. Finally, while increased prior knowledge was associated with higher single- and multiple-text comprehension, there was no interaction between prior knowledge and emotions on multiple-text comprehension. Overall, this study found that (1) learners often express more than one emotion during GBLE activities, (2) emotions expressed while learning with a GBLE shift across different activities, and (3) emotions are related to demonstrated comprehension, but the type of activity influences this relationship. Results from this study provide implications for how emotions can be examined as learners engage in GBLE activities as well as the design of GBLEs to support learners' emotions accounting for different activity demands to increase comprehension of single and multiple texts.

Introduction

As game-based learning environments (GBLEs) become increasingly prevalent in classrooms, it is critical to examine whether learners' prior knowledge and emotions impact their capacity to process information necessary for accurately comprehending text. As most learners cannot accurately monitor and regulate cognitive processes (National Academies of Sciences, Engineering, and Medicine [NASEM], 2018), GBLEs are designed with features (e.g., autonomy, interactions with nonplayer characters) that support conceptual understanding of complex scientific topics and maintain learners' cognitive and emotional engagement (Loderer et al., 2020; Plass et al., 2020; Sabourin & Lester, 2014; Taub et al., 2020). Compared to conventional education settings, GBLEs provide learners with opportunities to regulate their learning using affordances from GBLE features such as choice (i.e., autonomy to self-regulate their learning), nonlinear access to instructional content (e.g., content

CONTACT Daryn A. Dever ddever@knights.ucf.edu University of Central Florida, 4000 Central Florida Blvd, Orlando, FL 32816

exploration), and linked representations (e.g., texts providing different information on the same topic; National Academies of Sciences, Engineering and Medicine, 2018). For example, GBLEs including Operation ARA and Crystal Island have been designed with embedded virtual texts that learners must comprehend by (1) monitoring and effectively regulating cognitive processes and emotions while reading and (2) initiating cognitive strategies to process information from both single and multiple texts within the GBLE (Forsyth et al., 2020; Halpern et al., 2012; Millis & Halpern, 2014; Rowe et al., 2011). By deploying these processes and strategies during reading instructional materials, learners achieve greater accurate reading comprehension performance (Bohn-Gettler, 2019). However, the emotions experienced as a by-product of encountering complex instructional content, such as scientific concepts, and completing activities incorporated within the GBLE can influence the extent to which learners accurately comprehend information (Cloude et al., 2020; D'Mello & Graesser, 2012; Forgas, 2002). Therefore, it is important to specifically study the role of texts on emotions, individual- (e.g., prior knowledge), and activity-based factors – i.e., reproductive processing for single-text comprehension vs constructive processing for multiple-text integration – to understand how learners process information and its relation to reading comprehension within GBLEs.

Emotions

Reading comprehension models and literature examining the relationship of emotions and information processing strategies (e.g., PET framework; Bohn-Gettler, 2019) explain that emotions affect learners' cognition. Emotions are defined as states reflecting individuals' judgments of their progress toward goals to coordinate current and future behavior (Hudlicka, 2017) and described across dimensions such as valence and discrete vs non-discrete (Azevedo et al., 2018; D'Mello & Graesser, 2015; D'Mello et al., 2018; Harley et al., 2016; Pekrun & Linnenbrink-Garcia, 2012). Valence traditionally characterizes emotions as either pleasant/positive (e.g., joy) or unpleasant/negative (e.g., anger) experiences. Empirical studies find that positive and negative emotions can be both beneficial and detrimental to cognitive engagement and information processing, ultimately impacting learning (Cloude et al., 2020; D'Mello & Graesser, 2012; Forgas, 2002; Pekrun & Linnenbrink-Garcia, 2012). Discrete vs non-discrete emotions refer to whether emotions are experienced with or without the presence of another emotion (e.g., surprise vs surprise-anger; Azevedo et al., 2018).

Past work asserts emotion valence (i.e., positive vs negative) and individual-based factors (e.g., prior knowledge) impact how learners select required processing strategies (e.g., reproductive vs constructive processing) where this interaction is further moderated by activity-based factors, or the nature and demands of the task required to successfully complete different activities (i.e., does the task require learners to recall information from one text or integrate information from multiple texts?; Bohn-Gettler, 2019). Therefore, to understand how emotions impact learners' cognition and subsequent performance on tasks, it is critical for researchers to account for how (1) emotions are characterized; (2) emotions are contextualized in terms of the environment and activities in which they are experienced; and (3) learner's individual differences, such as prior knowledge, influence emotions. In understanding the role of emotions, activity, and prior knowledge on cognitive processing, research can provide direction for designing adaptive GBLEs capable of scaffolding learners' information processing and enhancing their reading comprehension.

Emotions and activity-based factors

GBLEs are designed to induce emotions using narrative and other features that cognitively and emotionally engage learners as they interact with instructional materials (Plass et al., 2020). As learners read during game-based learning, they must regulate their cognition and emotions to successfully engage in the processing (e.g., constructive processing) required for accurate text comprehension and increased task performance (Azevedo et al., 2018; McRae & Gross, 2020; Taub et al., 2018). Yet, learners' ability to effectively use cognitive processes to comprehend information during activities and tasks within the GBLE is dependent upon learners' emotions (Azevedo et al., 2018; D'Mello & Graesser, 2012; Pekrun & Linnenbrink-Garcia, 2012). For example, experiencing confusion while reading text within a GBLE can impact how learners engage in cognitive processing during the activity of reading, manage their cognitive resources (e.g., emotions require processing, limiting resources available to other tasks; cognitive load; Sweller, 1988), and select strategies (e.g., rereading, selecting texts, information-processing strategies) to resolve confusion and reach comprehension (D'Mello & Graesser, 2012, 2015). In employing accurate cognitive processes that result in increased reading comprehension, such as utilizing reproductive processing for a recall task, previously expressed confusion may transition to joy (D'Mello & Graesser, 2012). This transition of emotions during learning with a GBLE can significantly impact the extent to which learners can demonstrate understanding during performance tasks (Di Leo et al., 2019).

While previous work has highlighted the importance of emotions during learning with GBLEs (Azevedo et al., 2018; D'Mello & Graesser, 2012; Pekrun & Linnenbrink-Garcia, 2012; Plass et al., 2020), it is essential to also consider the moderating effect of the type of cognitive processing that is required for specific activities within GBLEs to increase reading comprehension. During reading activities, integrating and validating information from text is described as either *reproductive* or *constructive* processing (Bohn-Gettler, 2019). *Reproductive information processing* occurs when the task or activity does not require learners to engage in reorganizing their mental representations of constructs, such as relaying (i.e., recalling) factual information directly from the text to demonstrate single-text comprehension (Bohn-Gettler, 2019). Conversely, *constructive information processing* requires learners to transform their mental representations, fill in missing information, problem-solve, etc. (Bohn-Gettler, 2019). Constructive processing may occur during the integration of information from multiple sources or texts where learners are required to make connections between different sources of information to successfully complete a task (Rouet et al., 2019).

Not only does the type of processing required by learning activities influence comprehension but, as the PET framework suggests, it is possible that these processes interact with learners' emotions as well (Bohn-Gettler, 2019; Van den Broek & Helder, 2017). A study by Trevors et al. (2016) examined how emotions were related to reading comprehension and cognitive processing. This study showed that negative emotions were associated with decreased reading comprehension performance when the task required learners to engage in reproductive processing to recall factual information directly from the text to successfully complete the posttest. Such results emphasized a dissociation between task requirements and selected information-processing strategies and their differing effects on emotions and comprehension performance. Similar results were found in a paper by Mills et al. (2017) that examined how emotions influenced the extent to which learners were able to comprehend expository text. The authors exposed learners to either a happy or sad film prior to the reading activity and found that learners who read text inducing negative emotions, compared to learners who read text inducing positive emotions, had greater reading comprehension demonstrated on questions that required constructive processing (i.e., deep-reasoning), but not questions requiring lower-level cognitive processes (e.g., information recall). This suggests, both the type of emotion elicited during reading and the activity-based factors (i.e., constructive vs reproductive processing) resulted in varying levels of effectiveness of learners' cognitive processing. Specifically, this study showed that negative emotions (i.e., sadness) were essential to constructive processing but not reproductive processing.

Another study by Scrimin and Mason (2015) examined how positively-, neutrally-, and negatively-induced emotions were related to learners' eye movements and text comprehension. This study induced emotions by having learners watch a video clip, then read scientific text as their eye movements were recorded. In contrast to the previous studies, Scrimin and Mason (2015) found that learners who had induced positive emotions prior to reading scientific text had greater fixation durations on text while re-reading, interpreted as indicating greater in-depth and purposeful processing, and demonstrated greater text comprehension on a pretest including both factual and transfer questions. Results from Scrimin and Mason (2015) study, in comparison to the previously mentioned studies, highlight mixed findings in literature on how emotions are related to activity-based factors and subsequent reading comprehension.

Emotions and individual-based factors

In addition to examining the cognitive processing required during the activity for successfully completing a comprehension task, it is essential to consider individual-based factors. Individual differences (e.g., prior knowledge) influence learners' ability to accurately and extensively engage in cognitive processing and the type of cognitive processes employed (McCardle & Hadwin, 2015; Trevors et al., 2017). Prior knowledge must be activated for learners to recall information from a single text via reproductive processing and integrate information from multiple texts and prior knowledge during constructive processing (Bohn-Gettler, 2019). During single-text comprehension, it is essential for the learner to integrate information on a smaller scale such as across sentences and paragraphs (Kintsch & van Dijk, 1978). However, multiple texts refer to information that is found in multiple sources and related to a common topic (e.g., microbiology). As such, comprehending information across multiple texts is influenced by the relationship between sources, including the redundancy or inconsistency of information, the source from which the information comes, and the way in which learners employ strategies when they encounter multiple texts compared to a single text (e.g., the order in which texts are read; Rouet et al., 2019). To successfully incorporate information from both single and multiple texts, learners are required to activate their prior knowledge (Braasch et al., 2013; Bråten et al., 2014; Rouet et al., 2019).

Previous studies have found that both the level of prior knowledge (i.e., high prior knowledge vs low prior knowledge) and processing required to comprehend the text (e.g., recalling information vs constructing a new mental model from information) influences how learners interact with text. Learners with high prior knowledge are better able to integrate relevant information in their mental model and focus on activating this prior knowledge compared to learners with little to no prior knowledge about the content (Jarodzka & Brand-Gruwel, 2017). A study by Van Moort et al. (2018) examined how reading comprehension was influenced by learners' ability to evaluate and monitor their understanding of information using content from the text and prior knowledge. Reading comprehension was measured using time on text which contained information conflicting the provided text and prior knowledge. Conflicting information relative to text would be two contradictory sentences with one preceding the other, while conflicting information relative to prior knowledge would be information that contradicted learners' prior knowledge. Results indicated that both evaluating and monitoring text- and knowledge-based information influenced learners' ability to select and process correct information to achieve reading comprehension. Specifically, evaluating and monitoring knowledge-based information showed longer time spent reading text, suggesting that learners require time to comprehend information that contradicts their prior knowledge (Van Moort et al., 2018). Overall, prior knowledge is essential to cognitive processing since the extent of learners' prior knowledge about the domain can impact which cognitive processes are utilized and the accuracy of information processed.

Research has studied the interaction between emotions and cognitive processing using refutation text – i.e., text that contains information that intentionally presents a misconception to the reader and then refutes the misconception (Danielson et al., 2016; Trevors et al., 2017). A study by Trevors and Kendeou (2020) examined the influence of both positive and negative induced emotions on refutation texts about vaccines that required learners to activate their prior knowledge during the task. This study found that regardless of emotions, emotions induced by the content of the refutation text, overall knowledge on the subject increased. When accounting for the level of prior knowledge a learner has, a study by Zaccoletti et al. (2019) found that learners who were proficient in integrating information with their prior knowledge demonstrated higher reading comprehension outcomes regardless of whether they experienced negative emotions. Yet, another study by Storbeck and Clore (2005) supports emotions as closely related to cognitive processing requiring prior knowledge activation, such as constructive processing. This study by Storbeck and Clore (2005) examined how induced emotions are associated with the accurate activation of prior knowledge where learners induced with negative emotions were less likely to inaccurately activate prior knowledge when recalling information from a task whereas learners with positively induced emotions were more likely to demonstrate a false memory effect.

These studies, similar to literature regarding activity-based factors, highlight the mixed results found when examining the relationships between emotions, prior knowledge, and cognitive processing. These mixed findings for both activity- and individual-based factors may be due to the several limitations within current research on the methodology and characterization of the investigated emotions.

Theoretical framework

The goal of this study is to examine the extent to which learners' emotions, individual-based factors (i.e., prior knowledge), and activity-based factors (i.e., single text comprehension vs multiple-text integration) within a GBLE were related to performance on in-game and posttest measures of reading comprehension. We grounded our work in Bohn-Gettler's (2019) Process, Emotion, Task (PET) framework which describes relationships between reading comprehension, activity-based factors of single- and multiple-text comprehension, individual-based factors, and emotions. Specifically, we used PET (Bohn-Gettler, 2019) to investigate both how shifts in emotions between activities (i.e., reading and in-game assessment) and prior knowledge as an individual-based factor relate to reading comprehension performance during and after learning with a GBLE. The PET framework converges multiple comprehension concepts, theories, and models to examine text-, individual-, and activity-based factors on information processing, emotions, and comprehension (Bohn-Gettler, 2019). This framework proposes several hypotheses supported by comprehension literature and emotion theory.

For the objectives of this paper, we review Hypothesis 7 which states text-, individual-, and activity-based variables interact with emotions to mediate and moderate comprehension. Individual-based factors, according to Hypothesis 7 (Bohn-Gettler, 2019), refers to the individual differences of learners that can influence emotions, such as prior knowledge, working memory capabilities, personality, and reading skill. Activity-based factors are defined as the cognitive processes learners employ that are essential to completing the tasks or activities (Bohn-Gettler, 2019). For example, single-text comprehension requiring reproductive processing is an activity-based factor, as is, conversely, multiple-text integration requiring constructive processing. Throughout this paper, we use the PET framework (Bohn-Gettler, 2019) to guide our discussion on the relationship between emotions and single- and multiple-text reading comprehension captured during both reading and learning activities.

Current study

We address the mixed findings of the reviewed literature by examining the extent to which emotions relate to learners' individual- and activity-based factors. As we have described, previous work has examined how emotions influence cognitive processing during reading and subsequent performance on reading comprehension tasks. However, this work has not considered both factors together or when a learner's emotions occur. This study uses the future directions set forth by Bohn-Gettler's (2019) PET framework calling for the systematic examination of specific comprehension processes in relation to particular emotions and features of the learning activity. Specifically, this study examines Crystal Island, a GBLE focused on increasing knowledge about microbiology knowledge. We analyze learners' expressed emotions during reading, how these emotions transition from reading to completing an in-game assessment, and how prior knowledge and expressed emotions during these activities relate to learners' in- and postgame reading comprehension. We aim to address major gaps in combining GBLE and comprehension literature to advance the field of educational, cognitive, psychological, affective sciences, and discourse processes. Our research questions and hypotheses are discussed below.

Research Question 1: How are emotions expressed as learners engage in different activities within a GBLE? This question explores which emotions are present as learners read an individual scientific text about microbiology within a GBLE and complete in-game comprehension assessments. Specifically, this question examines how joy, anger, confusion, and frustration are expressed by each learner as they interact with both activities. We hypothesize that there will be evidence of multiple emotions expressed both during reading and assessment activities, and that these emotions change across activities. We do not hypothesize a direction in which the characteristics of emotions, such as emotion valence (e.g., joy vs anger), will differ between tasks as this is an exploratory analysis of emotion expression data and existing literature does not align (Mills et al., 2017; Scrimin & Mason, 2015; Trevors et al., 2016).

Research Question 2: Do emotions differ as a function of the activity during learning with a GBLE? This question addresses how emotions are expressed across learning activities, reading and in-game comprehension assessments. According to Bohn-Gettler's (2019) PET framework, learners' emotions and cognitive processing will change depending on the task being completed. Because of this, it is critical to understand if emotions expressed are different between interrelated activities. We hypothesize that reading and in-game assessment, a measurement of single-text reproductive processing, instances will differ in the emotions expressed during each activity. We do not hypothesize a directional relationship as the literature has not been thoroughly expanded to include all constructs (Bohn-Gettler, 2019), but use this question to explore how emotions relate to different activities as learners interact with a GBLE.

Research Question 3: How do learners' expressed emotions relate to reading comprehension performance? This question examines how expressed emotions during different in-game activities, reading vs in-game assessments, were related to learners' successful completion of in-game and posttest comprehension measures. We hypothesize that emotions during reading and in-game assessment instances will be related to in- and postgame reading comprehension performance and that an increase in the number of emotions that learners experience will have a positive relationship with reading comprehension performance. This hypothesis is supported by previous literature on emotions and learners' ability to engage in cognitive processes required for these activities where learners will experience multiple emotions within a single activity as emotions are regulated for increased comprehension, indicating that a greater number of emotions expressed during a single activity instance will be associated with greater reading comprehension (Azevedo et al., 2018; D'Mello & Graesser, 2012, 2015; Pekrun & Linnenbrink-Garcia, 2012; Taub et al., 2018).

Research Question 4: To what extent are there relationships between prior knowledge, emotions, and reading comprehension performance? The goal of this question is to examine the relationship between prior knowledge, an individual-based factor, emotions expressed during reading and in-game assessments, and learners' posttest comprehension performance. Bohn-Gettler's (2019) PET framework hypothesizes that learner-based factors, such as prior knowledge, interact with emotions. This is supported by previous research and models (Azevedo et al., 2018; Kintsch & van Dijk, 1978; Mayer, 2019) where cognitive processes underlying comprehension require learners to integrate new information with their prior knowledge. As such, learners' level of prior knowledge affects their ability to identify the task, select relevant information from the environment, engage in both reproductive and constructive processing for in-game assessments which require recall and posttest measures that measure multiple-text comprehension. We, therefore, hypothesize that learners with higher prior knowledge will have greater reading comprehension performance both on in-game assessments and the posttest comprehension measure than learners with lower prior knowledge. Further, we hypothesize that prior knowledge will interact with emotions, but we do not hypothesize a directional relationship as literature has not fully explored the relationship between emotions and individual- and activity-based factors (Bohn-Gettler, 2019).

Methods

Participants

A total of 105 participants were recruited from three large North American universities. For this paper, a subset of 57 undergraduate students (M_{Age} = 20.11; SD_{Age} = 1.55; 65% female) were used for this study as they met the inclusion criteria: participants had complete facial expression, log-file, and performance data and were assigned to a condition in which they had full control over their actions in the learning environment. Of the sample, the majority identified as White (74%; n = 42), while the remaining identified as Asian (11%; n = 6), Hispanic (9%; n = 5), African American (4%; n = 2), American Indian (2%; n = 1), and other (2%; n = 1). The majority of students also reported rarely playing video games (42%; n = 24), while 23% reported occasionally playing video games (n = 13). Specifically, 68% reported that they played 0–2 hours of video games per week (n = 39), while 14% reported 3–5 hours per week (n = 8), 7% reported 5–10 hours per week (n = 4), and 10% reported 10–20 hours per week (n = 6). IRB approval was received prior to recruitment and data collection.

Crystal Island environment

Crystal Island (Rowe et al., 2011) is a narrative-based GBLE designed to teach learners about microbiological pathology – i.e., the characteristics of various pathogens. Crystal Island and associated tasks (i.e., pre-/posttests) were developed with the help of a subject matter expert in microbiology and designed to align with the Standard Course of Study Essential Standards for Eighth-Grade Microbiology (McQuiggan et al., 2008). To complete the game, learners must investigate a mysterious pathogen plaguing a research camp on a remote, tropical island. By adopting the role as a Center for Disease Control (CDC) agent who searches and gathers clues used to generate hypotheses about the source of the pathogen, learners can experimentally test their hypotheses to devise an effective treatment plan for the infected researchers. Specifically, learners are required to navigate through a research camp and converse with nonplayer characters (NPCs), as well as read books and research articles to learn about the symptoms of the illness and different pathogens. Once a general understanding of pathology is obtained, learners are instructed to formulate hypotheses about which pathogen may be infecting the researchers based on their symptomatology (e.g., fatigue, cough). Through conversing with NPCs such as the head chef and other members of the research team, learners gather clues about the source of the illness (e.g., sick researchers ate bread and eggs, whereas healthy researchers ate apples for breakfast) and test food items that could reveal the source of the pathogen. Learners insert the food items into a scanner where items are then tested for specific diseases, viruses, and/or bacteria. Upon locating the source of the illness, the learner can pinpoint which pathogen has infected the researchers and provide an effective treatment solution based on their understanding of pathology. To complete the game, learners must provide a correct source of the pathogen and plan of treatment for the sick researchers.

In Crystal Island, learners read a maximum of 21 texts, displayed as books and research articles, placed throughout the environment to obtain information needed to solve the mystery and successfully complete the posttest. These texts contain complex topics that learners must read to complete in-game assessments (see Figure 1). Each assessment corresponds to a text which holds all information needed and varies in the number of multiple-choice cells that require information, ranging from three to eight cells. For example, an assessment may ask the learner to report the characteristics of a tapeworm and ways to prevent contracting a tapeworm. Once a learner selects a cell to edit, a drop-down menu of four possible answers will appear with only one correct answer. Once an answer is selected, the learner will fill in all cells and submit the assessment. Upon submission, the learner is scored based on the number of correct and incorrect answers. If a learner does not submit all cells correctly, the learner receives feedback from the system which highlights the incorrect answer(s). The learner then has the option to either resubmit an additional two times or continue with the game.

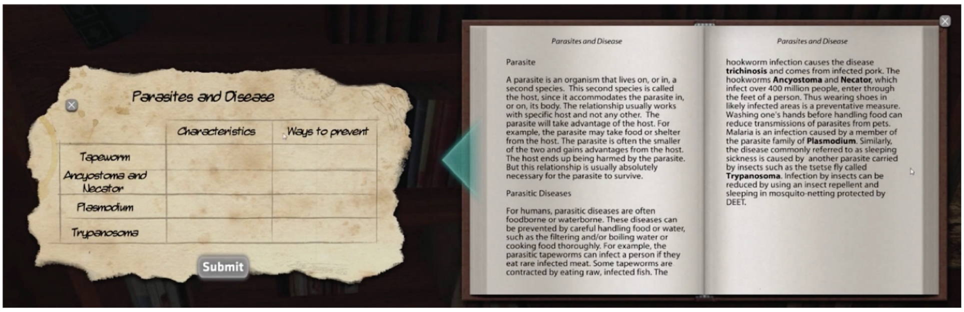

Figure 1. Examples of in-game assessment (left) and text (right) in Crystal Island.

Apparatus

Logfiles

To capture learners' actions during game-based learning with Crystal Island, we collected online, time-stamped trace data. These data revealed when, how often, and how long learners interacted with elements and tools, such as books, research articles, and in-game assessments, during learning with Crystal Island.

Eye tracker

To record eye movements during game-based learning with Crystal Island, an SMI EYERED 250 eye tracker (SMI Experiment Center, 2014) detected learners' pupil and fovea location using infrared light. Specifically, we used a nine-point calibration to establish the highest accuracy and precision possible, and the eye-tracker was configured to capture eye movements on the screen at a sampling rate of 30 Hz, recording relatively small eye movements at an offset of 0.05 mm. Eye-gaze data were postprocessed using iMotions software (iMotions Attention Tool (Version 6.0) [Computer software], 2016) to establish areas of interest (AOIs) which were used to determine the amount of time learners spent fixating on in-game assessments, reading specific books and research articles, etc. Eye-gaze fixations were operationalized as learners looking at an AOI with a relatively still gaze for at least 250 ms (Salvucci & Goldberg, 2000).

Facial expressions of emotions

Facial expressions of emotions were automatically recorded and coded during game-based learning with FACET, a video-based facial expression tracking system (Taub et al., 2019), implemented using iMotion software. Specifically, this software captures facial features at a sampling rate of 30 Hz that are then algorithmically classified according to the Facial Action Coding System (FACS; Ekman et al., 2002) using a facial-detection framework: (1) image input, (2) feature detection, and (3) feature classification. This framework was used to determine an evidence score that reflects the presence of an affective state occurring at a particular video frame, based low-level, muscle contraction facial expression features called action units (AUs), such as furrowing the eyebrows (i.e., AU4). The FACET tracking system classified learners' facial expressions during game-based learning, providing evidence scores of the expression of emotions in real-time for 20 AUs and 10 emotions (e.g., anger, surprise, frustration, joy, confusion, fear, disgust, sadness, and contempt, neutral). Evidence scores represent the probability an expert human coder would categorize a given frame of a face as reflecting the intended AU or emotion (iMotions A/S, 2016).

Commercially available facial expression detection algorithms, such as FACET, have recently begun being validated and found to accurately classify prototypical emotions (e.g., Stöckli et al., 2018). However, these algorithms face the same criticisms as FACS. Primarily, these algorithms categorize expressions, not necessarily emotions, requiring additional inference and contextualization steps to classify emotions, especially non-prototypical emotions. Some argue this gap can be addressed with hybrid models (e.g., neural networks) that would combine empirical data (e.g., FACET evidence scores) and theoretical constraints (e.g., number of feedback loops, number of connections, etc. as defined by appraisal models; Mortillaro et al., 2015). We argue that less sophisticated aggregation methods can be used for non-discrete emotions. As such, we used these emotion evidence scores to create emotion groups (Research Questions 1, 3, and 4) as well as measures of the likelihood of the presence of an emotion (Research Question 2; see *Coding and Scoring*).

Experimental procedure

Participants entered the research laboratory, and upon giving informed consent, they were randomly assigned to one of three conditions.[1] After participants were instrumented and successfully calibrated to an electro-dermal bracelet, eye tracker, and facial expressions of emotions software, participants were instructed to complete a series of demographics questionnaires gauging age, race, major, and experience with playing video games, self-report measures capturing emotions (Achievement Emotions Questionnaire, AEQ; Pekrun et al., 2006), motivation (Intrinsic Motivation Inventory, IMI; Ryan, 1982), and goal orientation (Achievement Goal Orientation, AGO; Elliot & Murayama, 2008). This current study did not use information obtained from these self-reports. A 21-item, multiple choice pretest, administered online and developed by a subject matter expert in microbiology, quantified learners' level of understanding about microbiological pathology. Afterward, participants began learning with Crystal Island. At the beginning of the game, participants encountered a tutorial which demonstrated how to use the various tools in Crystal Island and illustrated how to navigate the environment effectively. Participants were required to provide a correct pathogen source and treatment solution to the camp, and in doing so, completed the game. Afterward, participants were instructed to complete the same self-report measures and a 21-item, multiple choice posttest administered online that was similar, but not identical, to the content in the pretest to avoid practice effects.

Coding and scoring

Reading and in-game assessment instances

Eye-gaze behaviors on AOIs were used to determine when a participant was reading a research article or book as well as completing the in-game assessment. We chose not to use log-file data because while a book might have been open on screen, it was not indicative of reading or working on the assessment. For this study, we defined an instance as the total time fixating on the in-game assessment or book until the first assessment submission. We were only interested in the first instances prior to the first submission of an in-game assessment as feedback was provided after submission which could have caused reactivity to the approach of rereading or subsequent in-game assessment attempts (e.g., gaming the system).

In-game assessment and posttest comprehension measures

The in-game assessments were displayed as a series of multiple-choice questions in a matrix (Rowe et al., 2011). These in-game assessments were coupled with both books and research articles and addressed information contained within the coupled text. As in-game assessments required participants to recall information directly mentioned in the text (see Table 1), in-game assessments measure participants' single-text comprehension, requiring participants to engage in reproductive processing. Further, in-game assessments did not require participants to depend on prior knowledge or integrate any information from previously encountered scientific text within Crystal Island. To quantify single-text comprehension, log files were used to determine how many edits after their original choice a participant made in any cell of the in-game assessment. For example, if a participant decided to change the "prevention method of a tapeworm" from "antibiotic" to "wear shoes," this was considered one additional edit for that assessment. The more additional edits a participant had, the less they

Table 1. Processing strategy associated with comprehension measurements.

Measurement	Processing Strategy Required	Reasoning	Example
In-game Assessments	Reproductive	Required participants to recall information directly from text that was just read.	Recalling the prevention method of a tapeworm (e.g., wearing shoes) directly from the text.
Posttest	Constructive	Required participants to integrate pieces of information from multiple texts read throughout the game. Factual and applied questions are addressed throughout multiple texts.	Answers to factual questions – e.g., *Which of the following can be passed directly from person to person?* – can be found throughout multiple sources.

demonstrated in-game comprehension of the material as participants were either guessing, did not fully read the text, misunderstood information, or attempted to game the system. Due to the nature of the assessment requiring a canned response (i.e., multiple choice), we assumed that edits did not reflect optimizing or elaborating on information but rather changing their comprehension.

Multiple-text comprehension was measured using the posttest which examined participants' comprehension performance over the entire game as participants were required to read and integrate information from all texts and use prior knowledge to successfully complete the posttest (see Table 1). The posttest, as well as the pretest, contained both factual questions ($n = 12$; e.g., "What is the smallest type of living organism?") requiring recall, and applied questions ($n = 9$; e.g., "A man is administered to a hospital experiencing the following symptoms: Muscle Paralysis, Vomiting, Nausea, and Stomach Cramps. What is the most likely cause of his illness?") that required participants to integrate information from multiple texts in Crystal Island and transfer information to other scenarios. Each question had four multiple-choice options with a single correct answer. Participants could only select a single solution. If the participant was correct, they would obtain one point for that question. If the participant answered incorrectly, they did not receive any points for that question. We used the participant's final posttest score as a measure of multiple-text comprehension. Initially, the pre- and posttests contained 21 questions, however, one question was excluded from analyses as the item contained conflicting information from that represented in Crystal Island, resulting in a total of 20 possible points. Specifically, a book provided information that bacteria can reproduce sexually or asexually. The posttest item did not give an option for both sexual and asexual reproduction to be associated with bacteria, so this question was excluded along with the corresponding pretest measure. Posttest scores were calculated out of 20 possible correct answers. Posttest scores were normally distributed, ranged from 40% (eight correct responses) to 100% (20 correct responses), had an average of 72.14% (14.4 correct responses), and a median of 75% (15 correct responses).

Prior knowledge measure

We used participants' performance on the pretest to determine the level of prior knowledge before interacting with Crystal Island. The pretest was nearly identical to the posttest where the number of factual and applied questions, the topics covered, the method in which learners reported their answers (i.e., four-choice multiple-choice answers), and how the posttest scores were calculated out of a possible total of 20 points were similar. The only difference was how the questions were worded. For example, where the pretest question would be "Viruses can have all of the following shapes except." the posttest equivalent would ask "Viruses are known to take which of the following shapes?" Pretest scores were normally distributed, ranged from 35% (seven correct responses) to 90% (18 correct responses), had an average of 59.92% (12.0 correct responses), and a median of 60% (12 correct responses).

Emotions

FACET provides evidence scores of video recordings of participant faces which describe how likely "expert human coders" would categorize the expression from the given frame as the intended emotion (i.e., anger, joy, frustration, and confusion; e.g., Taub et al., 2021). Scores range from −4 to +4, and are similar to a Z-score, centered around zero (which would reflect an even chance that the given frame would be coded as a neutral expression), in which positive scores indicate evidence that an emotion is present while a negative score indicates evidence that an emotion is absent. As we were only concerned with the presence of emotions, any negative scores were converted into zero values for our study. While FACET provides evidence scores for ten emotions, we chose only four emotions to examine for our study – joy, anger, confusion, and frustration. We examined these emotions for three major reasons: (1) joy, anger, confusion, and frustration have been shown to directly influence learning (Baker et al., 2010; Cloude et al., 2020; D'Mello & Graesser, 2015); (2) joy, anger, confusion, and frustration can be characterized on multiple dimensions (i.e., valence and discrete/non-discrete); and (3) our analysis used all combinations of emotions and, therefore, would provide too many groups with a large number of discrete emotions. Specifically expanding upon (3), as we used all combinations of our chosen emotions, with each additional emotion the number of combination groups grows exponentially (i.e., with four discrete emotions there are 16 combination groups but with five discrete emotions, the number of combination groups jumps to 32). Such an increase in groups would result in too many categories for our analysis given our sample size.

Discrete vs non-discrete emotions

It is important to note that the FACET algorithm classifies emotions based on FACS, a protocol built to analyze basic emotions as discrete states. Evidence scores generated by the FACET algorithm are not equipped to classify non-discrete emotions. To deal with this challenge, we classified joy, confusion, anger, and frustration based on their presence throughout a learning task. This was aligned with eye-tracking and log-file data to identify the period of time learners expressed emotions while actively reading a book or completing an in-game assessment. Specifically, we identified all combinations of these four emotions being expressed and classified as present via the FACET algorithm (16 groups; see Table 2). If a learner was reading a book and the algorithm classified the presence as "anger" by generating an evidence score of 1.5 and "confusion" using an evidence score of 2.0, the data suggested that the learner was expressing "anger-confusion" rather than solely "confusion" (as it was the higher evidence score). Our raw measurement of emotion expressions was captured and algorithmically classified on a frame-by-frame granularity according to FACET. However, we are analyzing our data at a higher time-based aggregation (as a matter of seconds) and, therefore, our classification of discrete vs non-discrete emotions follows that same granularity. That is, discrete emotion expression events (e.g., anger) are identified when only one emotion is

Table 2. Reading and in-game assessments by emotion presence frequencies.

Emotion Group	Reading Frequency	Reading Percentage	In-game Assessment Frequency	In-game Assessment Percentage
1. Joy	0	0.00%	0	0.00%
2. Anger	118	12.62%	172	18.40%
3. Confusion	0	0.00%	0	0.00%
4. Frustration	0	0.00%	0	0.00%
5. Joy-anger	14	1.50%	18	1.93%
6. Joy-confusion	0	0.00%	0	0.00%
7. Joy-frustration	0	0.00%	0	0.00%
8. Anger-confusion	115	12.30%	155	16.58%
9. Anger-frustration	0	0.00%	0	0.00%
10. Confusion-frustration	0	0.00%	0	0.00%
11. Joy-anger-confusion	0	0.00%	0	0.00%
12. Joy-anger-frustration	0	0.00%	0	0.00%
13. Joy-confusion-frustration	0	0.00%	0	0.00%
14. Anger-frustration-confusion	0	0.00%	0	0.00%
15.Joy-anger-confusion-frustration (ALL)	688	73.58%	590	63.10%
16. None	0	0.00%	0	0.00%

expressed within a reading or in-game assessment instance and non-discrete emotion events (e.g., all emotions, anger-confusion, etc.) are identified when more than one emotion is expressed within a single activity instance.

Data processing

Data processing was completed through a Python (Python Core Team, 2015) pipeline, which collected and cleaned the process data collected from log files for analyses. The statistical program R (R Core Team, 2017) was run to conduct the statistical analyses for all research questions. The Base R and "MASS" (Venables & Ripley, 2002) packages were used for data analyses. Reading and in-game assessment instances were defined using eye-tracking fixations on predefined AOIs. Emotion evidence scores during reading and in-game assessment instances were provided by iMotions Attention Tool (Version 6.0) [Computer software] (2016). All negative emotion scores (negative likelihood of a human coder rating an emotion as present) were replaced with scores of 0.

Results

Research question 1: how are emotions expressed as learners engage in different activities within a GBLE?

To examine which emotions were expressed across activities, we calculated the frequency of four discrete emotions (i.e., joy, anger, confusion, frustration) expressed, and their multiple combinations of non-discrete emotions (e.g., joy-anger, all, etc.; see *Coding and Scoring*), both during reading and completing in-game assessments (see Table 2). Expressed emotions could occur during any point within the activities, in which participants spent an average of 41.11 s (SD = 29.35) reading (n = 935) in game and an average of 13.46 s (SD = 9.18) completing in-game assessments (n = 935). Broadly, out of a possible 16 emotion groups that could be characterized as discrete and non-discrete, only four emotion groups were found to be present: anger, joy-anger, anger-confusion, and all. We found 73.58% (n = 688) of participants expressed all four emotions across all reading instances, and 63.10% (n = 590) of participants expressed at least one emotion across all in-game assessment instances. Interestingly, anger was the only emotion to be expressed independently during a learning activity where all other emotions (i.e., joy, confusion, frustration) occurred in the presence of another emotion. However, anger was also expressed in conjunction with confusion, in 12.30% (n = 115) of reading and 16.58% (n = 155) of in-game assessment instances. Anger was also expressed with joy in 1.50% (n = 14) and 1.93% (n = 18) of reading and in-game assessment instances, respectively.

Research question 2: do emotions differ as a function of the activity during learning with a GBLE?

To examine how joy, anger, confusion, and frustration differed between learning activities, we compared the differences of expressed emotions during reading instances and their subsequent in-game assessment instances. Using the mean differences of evidence scores between paired instances,

Table 3. Mean evidence scores by emotion and mean change between paired reading and in-game assessment instances.

Emotion	Reading Evidence Scores M (SD)	In-game Assessment Evidence Scores M (SD)	Paired Difference (In-game Assessment Evidence Scores – Reading Evidence Scores) M (SD)
Joy	0.44 (0.70)	0.43 (0.74)	−0.01 (0.34)
Anger	0.78 (0.71)	0.79 (0.76)	0.01 (0.25)
Confusion	0.56 (0.56)	0.58 (0.60)	0.02 (0.18)
Frustration	0.54 (0.58)	0.56 (0.63)	0.03 (0.21)

we conducted a Wilcoxon signed-rank test given the unequal group sizes and distribution of evidence scores. Table 3 reports the means and standard deviations of the evidence scores of an emotion's presence during the learning activities. Joy had higher evidence scores during reading than in-game assessment instances (V = 143,261, p < .05) whereas learners' expression evidence scores of anger (V = 149,808, p < .05), confusion (V = 134,306, p < .05), and frustration (V = 140,493, p < .05) were greater during in-game assessment instances than reading instances. Frustration showed the largest change in emotion evidence scores from reading to in-game assessments.

Research question 3: how do learners' expressed emotions relate to reading comprehension performance?

To understand how emotions expressed during learning with a GBLE were related to reading comprehension performance, we classified reading and in-game assessment instances by the emotions that were expressed during these activities – i.e., anger, joy-anger, anger-confusion, all (see Research Question 1). In other words, reading and in-game assessment instances were classified into one emotion group depending on the emotion(s) expressed during each activity. From this, we examined how expressed emotions changed across coupled reading and in-game assessment instances. Figure 2 illustrates how each instance within reading emotion groups transitioned to their in-game assessment emotion groups where the width of each bar is proportional to the number of instances (n = 935) originating from the reading emotion group and transitioning into the in-game assessment emotion

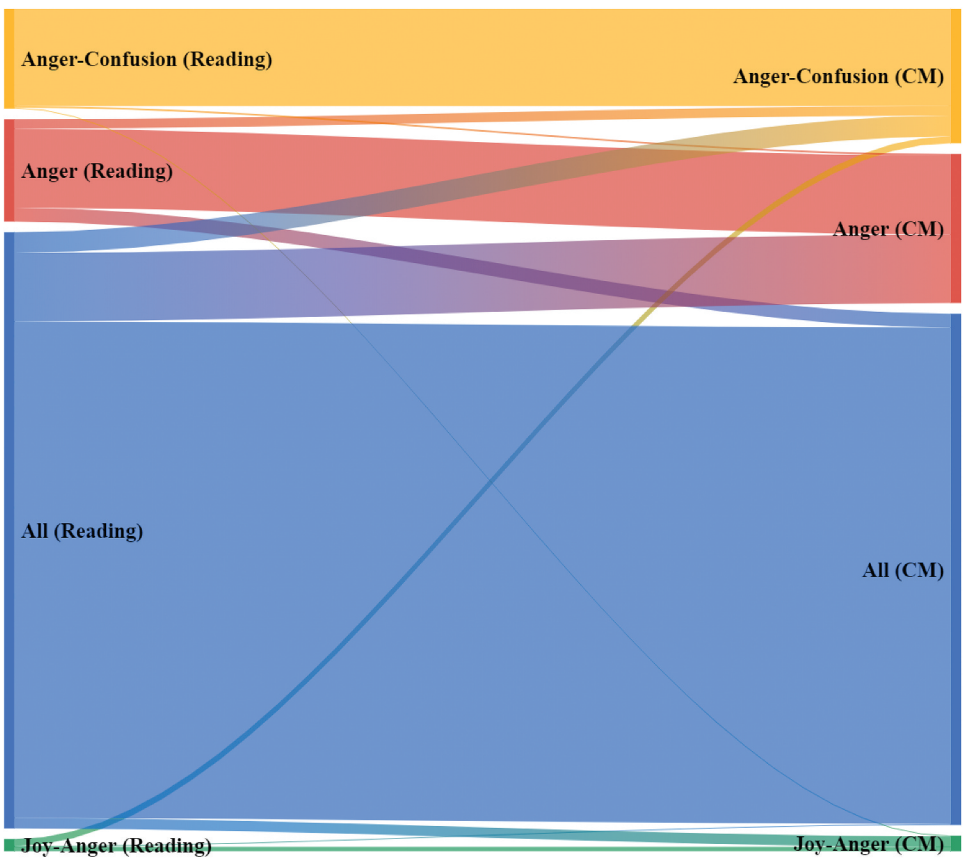

Figure 2. Sankey diagram of emotion group transitions from emotion groups during reading instances into emotion groups into in-game assessment instances.

groups. We see that most instances do not deviate from their original reading emotion group. That is, if an instance was classified as showing evidence scores of all emotions during reading, the majority of those instances would also show evidence scores of all emotions during the in-game assessment measure. It is interesting to note that approximately half of the reading instances for the “joy-anger” group (57.14%) transitioned to “anger-confusion” during the in-game assessment measure as this indicates that learners who experienced joy while reading transitioned to confusion as they were tested on their comprehension of the reading material. In addition, approximately half (47.93%) of the in-game assessment measure instances in the “anger” emotion group originated from the “all” emotions group classified during reading.

This research question further aims to understand how emotions expressed during learning with a GBLE are related to reading comprehension performance by examining how (1) emotions expressed during reading, classified into reading emotion groups (see Research Question 1) relate to in-game assessment performance measuring single-text comprehension; (2) emotions expressed during in-game assessments, classified into reading emotion groups (see Research Question 1) relate to in-game assessment performance measuring single-text comprehension; and (3) a combination of emotions groups expressed during reading and in-game assessments relate to posttest performance measuring multiple-text comprehension.

A Kruskal–Wallis test found significant differences in in-game performance between emotion groups during reading instances, $\chi^2(3) = 57.02$, $p < .01$ (see Table 4 for descriptive statistics). Post hoc pairwise comparisons using Dunn’s (1964) test with a Bonferroni correction for multiple tests (see Table 5) revealed that reading instances classified as containing “all” emotions demonstrated significantly lower in-game assessment comprehension than the other three emotion groups (i.e., “anger,” “joy-anger,” and “anger-confusion”; see Table 4). In sum, we found that reading instances in which there were evidence scores for all emotions present, participants demonstrated lower in-game comprehension of single texts compared to reading instances expressing evidence scores for “anger,” “joy and anger,” or “confusion and anger” emotion groups.

When classifying in-game assessment instances by emotion groups, we found significant differences of in-game assessment performance, $\chi^2(3) = 67.470$, $p < .01$ (see Table 4 for descriptive statistics). Pairwise comparisons (see Table 5) revealed that there were significant differences in performance between the instances with evidence scores for all emotions compared to “anger” instances and “anger-confusion” instances. Interestingly, unlike instances classified during reading, the instances classified as “joy-anger” while completing the in-game assessment measure revealed no significant difference in performance as instances with evidence scores for all emotions. Simply, we found that the in-game

Table 4. Means and standard deviations of additional edits associated with the emotion groups classified during reading and in-game assessments.

Emotion Group	Reading M (SD)	In-game Assessment M (SD)
Anger	0.99 (0.43)	0.96 (0.42)
Joy-anger	0.72 (0.45)	1.03 (0.34)
Anger-confusion	0.84 (0.39)	0.91 (0.46)
All	1.26 (0.57)	1.29 (0.58)

Table 5. Pairwise comparisons (z-scores) of additional edits by emotion group during reading and in-game assessment instances.

	All Emotions Group		Anger Emotions Group		Joy-anger Emotions Group	
Emotions Group	Reading	In-game Assessment	Reading	In-game Assessment	Reading	In-game Assessment
Anger	4.201**	6.250**				
Joy-anger	3.001*	1.434	1.386	−0.801		
Anger-confusion	6.296**	6.525**	1.646	0.427	−0.6216	0.987

*$p < 0.05$; **$p < 0.005$.

assessment performance was significantly lower between instances with evidence scores for all emotions compared to instances with anger or anger-confusion. That is, learners who had evidence scores for all emotions had less comprehension of scientific texts than learners who had either discrete emotion (i.e., "anger") or emotion-pair (e.g., "anger-confusion") evidence scores suggesting that not only the presence, but also the number of expressed emotions learners express during learning affects their information processing.

To understand how emotion relates to learners' multiple-text comprehension performance, we classified each participant by the emotion group to which the majority of their reading and in-game assessment instances belonged (see Research Question 1), removing any participants that had an equal number of instances between emotion groups (see Table 6 for descriptives). A Kruskal–Wallis test found no significant differences in performance on the posttest comprehension measure between emotion groups during reading or in-game assessments ($p > 0.05$).

Overall, results indicate that during activities requiring reproductive processing for single-text comprehension, there is an increased concentration of discrete emotions and emotion-pairs expressed compared to all emotions. Further analyses show that with a greater number of expressed emotions, there is a decrease in in-game assessment, but not posttest reading comprehension performance, signifying that the number of emotions expressed during learning activities can influence cognitive processing, demonstrated by single- and multiple-text comprehension measure performance.

Research question 4: to what extent are there relationships between prior knowledge, emotions, and reading comprehension performance?

A simple linear regression was used to predict in-game assessment performance from prior knowledge. Prior knowledge explained a significant amount of the variance in in-game assessment performance, $F(1,61) = 7.25$, $p < .01$, $R^2 = 0.11$. The regression coefficient ($\beta = -1.343$, 95%CI: -2.340, -0.346) indicated that the higher a participant performed on the pretest the better they did on in-game assessments of comprehension.

Next, we used a multiple linear regression model using the enter method to predict posttest reading comprehension from participants' (1) prior knowledge, (2) emotion groups categorized during reading, (3) emotion groups categorized during in-game assessments, and (4) interactions between prior knowledge and emotion groups (see Table 7 for correlation matrix of predictors). Due to the small group sizes for most emotion groups (see Table 6), we dichotomized emotion groups (for both reading and in-game assessment predictor variables) into "all" emotions (coded as 1) or "other" (coded as 0) which is a collection of the other emotion groups. In addition, we included the interaction of prior knowledge and the emotion groups in reference to Hypothesis 7 of Bohn-Gettler's (2019) PET

Table 6. Means and standard deviations of posttest scores by instance type (reading and in-game assessment) and emotion group.

Emotion Group	Posttest Scores by Reading Instances M (SD)	Frequency of Reading Groups	Posttest Scores by In-Game Assessment Instances M (SD)	Frequency of In-game Assessment Groups
Anger	0.78 (0.15)	6	0.69 (0.17)	10
Joy-anger	–	0	–	0
Anger-confusion	0.81 (0.14)	7	0.82 (0.13)	8
All	0.71 (0.14)	42	0.72 (0.13)	37

Table 7. Correlation matrix of covariates (r) used in a multivariable regression model of posttest performance.

	Pretest Scores	Reading Emotion Group
Reading emotion group	0.005	
In-game assessment emotion group	0.145	<0.001

Table 8. Parameter estimates, standard error, *t*-statistics, and *p* for a multivariable regression model of posttest performance.

	β	Std Error	*t*-value	*p*
(Intercept)	0.336	0.199	1.688	0.048*
Pretest scores	0.657	0.281	2.341	0.023*
Reading emotion group (0: other; 1: all)	−0.280	0.284	−0.984	0.330
In-game assessment emotion group (0: other; 1: all)	0.437	0.219	1.999	0.050*
Pretest scores * reading emotion group	0.510	0.482	1.058	0.300
Pretest Scores * in-game assessment emotion group	−0.778	0.415	−1.874	0.067

*Significance at $p < 0.05$.

framework, which suggests emotions and individual-based factors (e.g., prior knowledge) interact during information processing.

Our model explained a statistically significant amount of variance in posttest performance, $F(5,49) = 5.91$, $p < .01$, $R^2 = 0.31$. See Table 8 for coefficient estimates, standard error, *t*-statistics, and significance of all predictors. According to this model, pretest scores were a significant predictor for posttest reading comprehension scores ($\beta = 0.657$, 95%CI: 0.093, 0.736, $p = .023$), or simply, students that performed well on the pretest tended to perform well on the posttest reading comprehension measure. Emotions groups categorized during reading were not a significant predictor of posttest performance ($p = .330$), nor was the interaction between this variable and prior knowledge ($p = .300$). However, emotion groups categorized during in-game assessments were a significant predictor for performance ($\beta = 0.437$, 95%CI: 0.021,0.877, $p = .05$). This suggests that participants who showed evidence of expressing all emotions (joy, confusion, anger, and frustration) while completing the majority of their in-game assessments tended to outperform their peers on the posttest comprehension measure who only expressed either anger, joy-anger or anger-confusion. However, the interaction between this variable and prior knowledge was not a significant predictor of posttest performance ($p = .07$).

Discussion

The goal of this study was to examine the extent to which learners' emotions, individual-based factors (i.e., prior knowledge), and activity-based factors (i.e., single text comprehension vs multiple-text integration) within a GBLE were related to reading comprehension. Our research was designed to address how emotion expressions transition as a function of the activity and how this is related to single- and multiple-text comprehension. Specifically, we identified (1) which emotions were present during reading and in-game assessment instances, (2) how emotions changed from reading to in-game assessment activities, (3) the role of emotions on in-game and posttest performance, and (4) the relationship between prior knowledge and emotions on posttest scores.

The first research question identified which emotions learners expressed while engaging in reading and in-game assessments within Crystal Island. Our findings were partially consistent with our hypothesis where learners expressed multiple emotions during reading and in-game assessment instances. We identified several groups of emotions including "joy and anger," "anger and confusion," "anger," and "all" (i.e., anger, joy, confusion, frustration) emotion groups. Anger was the only discrete emotion expressed during reading and in-game assessments, emphasizing the need to consider multiple dimensionalities (i.e., valence, discrete vs non-discrete) when examining emotions and how several emotions interact with each other. In examining emotions that occur within the same task, we were able to assess which emotions were expressed by learners, how this may be a product of the task requirements and their resulting comprehension of information within GBLEs.

To further understand this relationship, the second research question examined whether emotions differed as a function of the activity during learning with a GBLE. In other words, we examined whether emotion expressions were different between reading and in-game assessments. Results were consistent with hypotheses where there was a significant difference in the presence of joy, anger, confusion, and frustration between reading and in-game assessment instances. We observed a pattern between joy (a positive emotion), confusion (classified as both positive and negative), as well as anger and frustration (negative emotions), where learners expressed joy more often during reading than in-game assessments. During in-game assessments, confusion, anger, and frustration were expressed significantly more than during reading, with frustration having the largest difference across activities. This finding suggests emotions fluctuate as a function of the activity which follows the PET framework describing learners' emotions as reliant on specific activities and their task requirements (i.e., reproductive vs constructive processing). These results identify a potential for assigning directionality to the PET framework which acknowledges that there is a relationship but does not expand on exactly *how* certain emotions interact with activities. In-game assessments require learners to recall information directly from the text which, from the results of this study, are associated with negative expressed emotions. This contrasts with reading activities which were associated with greater positive expressed emotions and requires learners both to memorize information directly from the text for reproductive processing and identify how information from one text relates to prior knowledge and previously read texts with constructive processing.

Our third research question sought to examine learners' comprehension and whether it was related to expressed emotions and task demands. These findings were partially consistent with our hypotheses, which predicted that expressed emotions would be related to increased in-game assessment and posttest comprehension performance (Bohn-Gettler, 2019). Results supported our hypothesis where learners who expressed multiple shifts in emotions (i.e., "all" emotion group) during reading and in-game assessments displayed lower single-text comprehension defined by their in-game assessment performance. A possible explanation may be a combination of the fluctuating emotions and the presence of multiple emotions. Cognitive resources are needed for learners to identify and regulate emotions while simultaneously selecting and utilizing cognitive and metacognitive strategies for reading comprehension (McCardle & Hadwin, 2015). As such, learners' ability to perform well on the in-game assessments and apply the constructive processing strategies needed for multiple-text reading comprehension may have been hindered by learners' constantly changing emotions.

However, performance on the posttest comprehension measure was not related to expressed emotions. This finding did not support our hypothesis and was not consistent with previous literature as well as the PET framework (Bohn-Gettler, 2019). A possible explanation could be that as the posttest comprehension measure requires constructive information processing, we need to examine how emotions also interact with learners' prior knowledge. Therefore, the fourth research question examined relationships between prior knowledge, emotions, and multiple-text comprehension measured by the posttest. Results found that prior knowledge was positively related to both single- and multiple-text comprehension performance, supporting our hypotheses and Bohn-Gettler's (2019) PET framework. This finding is also consistent with previous literature (Azevedo et al., 2018; Kintsch & van Dijk, 1978; Mayer, 2019) where cognitive processes underlying comprehension require learners to integrate their prior knowledge. This finding is interesting as in-game assessments only require direct recall from text, not the integration of prior knowledge. Further results showed that emotions expressed during reading and the interaction between prior knowledge and emotions expressed throughout learning did not relate to single- or multiple-text comprehension. However, when learners expressed "all" emotions during in-game assessment instances, they tended to demonstrate better multiple-text comprehension. These findings do not support our hypothesis as prior knowledge did not interact with emotions to influence reading comprehension performance. Rather, these constructs influenced comprehension performance independently. A possible explanation of these findings could be that prior knowledge is different than prior knowledge activation where measuring prior knowledge at the beginning of the task does not necessarily indicate a learners' ability to activate the prior knowledge during reading and, therefore, may not directly interact with the emotions experienced throughout learning.

Overall, findings suggest implications for understanding relationships between individual-based factors, emotions, activity-based factors, and reading comprehension as well as methods for measuring this relationship during game-based learning. First, emotions are an important construct to be measured on a gross scale to understand how activity-based factors interact with emotions during

GBLEs. Second, an increase in the number of emotions expressed within one instance of reading and in-game assessments hinder single-text comprehension, but a high number of emotions expressed within one instance of in-game assessments are associated with greater multiple-text comprehension. This reinforces the hypothesis of Bohn-Gettler's (2019) PET framework where emotions interact differently depending on the activity and its task demands. Third, we identify prior knowledge as essential for tasks requiring reproductive and constructive processing, an integral component for fully understanding how emotions relate to information processing and overall comprehension on posttests as well as how learners experience and express emotions before and during a learning task within a GBLE.

Limitations

Although the current study advances the reading comprehension and emotion literature, it is important to note the limitations of this study. First, the method in which this study operationally defined and measured single- vs multiple-text comprehension reflects the relative task demands across all activities in Crystal Island and is limited by the way information was collected. This study measured single-text comprehension by in-game assessment performance and multiple-text comprehension by posttest performance. While we acknowledge that prior knowledge activation and multiple-text integration may occur during a single text as learners integrate prior knowledge during reading, we evaluate this in comparison to learners reading across multiple texts, integrating information across multiple texts, and activating their prior knowledge required for the posttest. In addition, eye tracking was aggregated at a single-text level, instead of word-by-word. This, combined with the fact that the in-game assessment measure reflected reproductive processing, limited our ability to identify how learners processed information across the singular text instances.

Second, this study only analyzed four emotions, excluding emotions that also potentially influence learning and reading comprehension (e.g., boredom; Baker et al., 2010). In addition, emotions were identified using FACET (Littlewort et al., 2011), a facial recognition software which calculated the probability that emotions were present. Because evidence scores are provided on a frame-by-frame basis, it classifies emotions without considering context or previous emotional evidence scores. Using emotions found with FACET, reading and in-game assessment instances were classified into the groups that appeared throughout a single instance. That is, even if only a single frame had evidence scores for one of the chosen emotions, it was considered present. Future studies should consider new analytical approaches for creating more reliable representations of emotions with durations (e.g., fluctuations based on temporal dynamics contextualized to the learning task).

Third, our study classified learners into emotion groups depending on which emotion groups were most frequent over all instances for each individual. Classifying instances and learners into groups simplifies the emotions throughout reading and comprehension elements and ignores the natural within-subject differences any one individual has. Third, as groups were determined using frequency and not the likelihood of being present, instances were classified into emotion groups assuming the presence all carried equal weight in the instance. In other words, an instance classified as "anger-confusion" might erroneously suggest that a participant was feeling equal parts angry and confused during that instance whereas in reality, they could have just briefly felt anger with confusion being the only emotion present during the majority of the instance.

In addition to the limitations regarding classifying emotion groups, further limitations to the definition of in-game assessment performance are acknowledged. Instances of high performance were defined by lower additional in-game assessment edits and low performance had higher additional edits to the in-game assessment. While we justified this with the assumption that more edits would indicate the learner guessing or gaming the system, this may also be a by-product of learners' conscientiousness or motivation to complete the assessment to the best of their ability. While this is an important construct to consider, examining the motivation of the learner is outside the scope of this study. In addition, analyses that would indicate if conscientiousness plays a role during in-game assessments (i.e., correlations between additional edits and posttest scores) would not be available as the nature of in-game assessments and posttest comprehension measures (i.e., testing declarative knowledge vs a mix of declarative and applied knowledge respectively) is incongruent. Future studies may investigate the role of motivation as a learner-based factor on in-game assessment performance and overall comprehension.

Conclusion and future directions

Overall, by addressing how emotions are related to individual- and activity-based factors during reading within Crystal Island, our findings expand the current state of literature on how learners interact with different GBLE elements and their relationship with overall learning. Specifically, our study (1) introduced a new analytical approach for the identification of discrete and non-discrete emotions that were present during reading and in-game assessment instances, (2) showed changes in the emotions expressed across activities from reading to in-game assessment that require learners to select and utilize different processing strategies, (3) found the number of expressed emotions within a single instance were negatively associated with single-text comprehension, but positively associated with multiple-text comprehension, and (4) suggests prior knowledge is essential for increased performance on single- and multiple-text comprehension measures.

This study emphasizes the need to consider how the nature of activities and their associated cognitive processing demands within the GBLE as well as learners' prior knowledge are related to emotions and demonstrated reading comprehension. By examining the totality of the relationship between these constructs using Bohn-Gettler's (2019) PET framework, we are able to expand our understanding of how GBLEs (and potentially additional learning environments; e.g., simulation) can (1) both hinder and encourage reading comprehension, and (2) necessitate the regulation of learners' emotions during reading and the monitoring of their comprehension through in-game assessments. Our findings have implications for future research, design, and development of GBLEs that can adapt to learners' emotions depending on the activity, the demands of the task, and learners' prior knowledge. Future research on this subject should consider a broader range of individual-based factors, such as motivation or goal orientation, to better understand how these factors influence learners' reproductive and constructive processing for single- and multiple-text comprehension and emotions (including the role of emotion regulation strategies; McRae & Gross, 2020). In progressing research in this area, we would be able to develop more intelligent technologies capable of providing individualized training in regulating emotions accounting for individual- and activity-based factors to achieve higher reading comprehension.

Note

1. Participants were split into either full, partial, or no agency conditions that varied in the amount of afforded autonomy by restricting possible actions participants could take. For the purposes of this study, only data from participants in the no agency condition, which did not restrict any actions, were used.

Acknowledgments

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 authors and do not necessarily reflect the views of the Social Sciences and Humanities Research Council of Canada.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

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 P. A. Alexander, D. H. Schunk, & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed., pp. 254–270). Routledge.

Baker, R. S. J. D., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human Computer Studies*, *68*(4), 223–241. https://doi.org/10.1016/j.ijhcs.2009.12.003

Bohn-Gettler, C. M. (2019). Getting a grip: The PET framework for studying how reader emotions influence comprehension. *Discourse Processes*, *56*(5–6), 386–401. https://doi.org/10.1080/0163853X.2019.1611174

Braasch, J. L. G., Bråten, I., Strømsø, H. I., Anmarkrud, Ø., & Ferguson, L. E. (2013). Promoting secondary school students' evaluation of source features of multiple documents. *Contemporary Educational Psychology*, *38*(3), 180–195. https://doi.org/10.1016/j.cedpsych.2013.03.003

Bråten, I., Anmarkrud, Ø., Brandmo, C., & Strømsø, H. I. (2014). Developing and testing a model of direct and indirect relationships between individual differences, processing, and multiple-text comprehension. *Learning and Instruction*, *30*, 9–24. https://doi.org/10.1016/j.learninstruc.2013.11.002

Cloude, E. B., Wortha, F., Dever, D. A., & Azevedo, R. (2020). How do Emotions change during learning with an intelligent tutoring system? Metacognitive monitoring and performance with MetaTutor. In S. Denison., M. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proceedings of the 42nd Annual Conference of the Cognitive Science Society* (pp. 423–429). Cognitive Science Society.

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, *22*(2), 145–157. https://doi.org/10.1016/j.learninstruc.2011.10.001

D'Mello, S., & Graesser, A. (2015). Feeling, thinking, and computing with affect-aware learning technologies. In R. A. Calvo, S. D'Mello, J. Gratch, & A. Kappas (Eds.), *The Oxford handbook of affective computing* (pp. 419–434). Oxford University Press.

D'Mello, S., Kappas, A., & Gratch, J. (2018). The affective computing approach to affect measurement. *Emotion Review*, *10*(2), 174–183. https://doi.org/10.1177/1754073917696583

Danielson, R. W., Sinatra, G. M., & Kendeou, P. (2016). Augmenting the refutation text effect with analogies and graphics. *Discourse Processes*, *53*(5–6), 392–414. https://doi.org/10.1080/0163853X.2016.1166334

Di Leo, I., Musi, K. R., Singh, C., & Psaradellis, C. (2019). Curiosity confusion? Frustration! The role and sequencing of emotions during mathematics problem solving. *Contemporary Educational Psychology*, *58*, 121–137. https://doi.org/10.1016/j.cedpsych.2019.03.001

Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics*, *6*(3), 241–252. https://doi.org/10.1080/00401706.1964.10490181

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial action coding system*. Netwprk Information Research Corporation.

Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology*, *100*(3), 613–628. https://doi.org/10.1037/0022-0663.100.3.613

Forgas, J. P. (2002). Feeling and doing: Affective influences on interpersonal behavior. *Psychological Inquiry*, *13*(1), 1–28. https://doi.org/10.1207/S15327965PLI1301_01

Forsyth, C. M., Graesser, A., & Millis, K. (2020). Predicting learning in a multi-component serious game. *Technology, Knowledge and Learning*, *25*(2), 251–277. https://doi.org/10.1007/s10758-019-09421-w

Halpern, D. F., Millis, K., Graesser, A. C., Butler, H., Forsyth, C., & Cai, Z. (2012). Operation ARA: A computerized learning game that teaches critical thinking and scientific reasoning. *Thinking Skills and Creativity*, *7*(2), 93–100. https://doi.org/10.1016/j.tsc.2012.03.006

Harley, J. M., Lajoie, S. P., Frasson, C., & Hall, N. C. (2016). Developing emotion-aware, advanced learning technologies: A taxonomy of approaches and features. *International Journal of Artificial Intelligence in Education*, *27*(2), 268–297. https://doi.org/10.1007/s40593-016-0126-8

Hudlicka, E. (2017). Computational modeling of cognition-emotion interactions: Theoretical and practical relevance for behavioral healthcare. In M. Jeon (Ed.), *Emotions and affect in human factors and human-computer interaction* (pp. 383–436). Academic Press.

iMotions A/S. (2016). *Attention toll FACET module guide 130806*. Retrieved August 07, 2020, from https://imotions.com/guides/

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

Jarodzka, H., & Brand-Gruwel, S. (2017). Tracking the reading eye: Towards a model of real-world reading. *Journal of Computer Assisted Learning*, *33*(3), 193–201. https://doi.org/10.1111/jcal.12189

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. https://doi.org/10.1037/0033-295X.85.5.363

Littlewort, G., Whitehill, J., Wu, T., Fasel, I., Frank, M., Movellan, J., & Bartlett, M. (2011). The computer expression recognition toolbox (CERT). *Proceedings from 2011 I.E. International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011)*, (pp. 298–305). Santa Barbara, CA, USA. IEEE.

Loderer, K., Pekrun, R., & Lester, J. (2020). Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments. *Learning and Instruction*, *70,* 101162. https://doi.org/10.1016/j.learninstruc.2018.08.002

Mayer, R. E. (2019). Computer games in education. *Annual Review of Psychology*, *70*(1), 531–549. https://doi.org/10.1146/annurev-psych-010418-102744

McCardle, L., & Hadwin, A. F. (2015). Using multiple, contextualized data sources to measure learners' perceptions of their self-regulated learning. *Metacognition and Learning*, *10*(1), 43–75. https://doi.org/10.1007/s11409-014-9132-0

McQuiggan, S. W., Goth, J., Ha, E., Rowe, J. P., & Lester, J. C. (2008, June). Student note-taking in narrative-centered learning environments: Individual differences and learning effects. In B. P. Woolf, E. Aïmeur, R. Nkambou, & S. Lajoie (Eds.), *International Conference on Intelligent Tutoring Systems* (pp. 510–519). Springer.

McRae, K., & Gross, J. J. (2020). Emotion regulation. *Emotion*, *20*(1), 1–9. https://doi.org/10.1037/emo0000703

Millis, G. A. C., & Halpern, D. F. (2014). Operation ARA: A serious game that combines intelligent tutoring and learning principles to teach science. In V. A. Benassi, C. E. Overson, & C. M. Hakala (Eds.), *Applying science of learning in education: Infusing psychological science into the curriculum* (pp. 169–183). Society for the Teaching of Psychology.

Mills, C., Wu, J., & D'Mello, S. (2017). Being sad is not always bad: The influence of affect on expository text comprehension. *Discourse Processes*, *56*(2), 99–116. https://doi.org/10.1080/0163853X.2017.1381059

Mortillaro, M., Meuleman, B., & Scherer, K. R. (2015). Automated recognition of emotion appraisals. In J. Vallverdu (Ed.), *Handbook of research on synthesizing human emotion in intelligent systems and robotics* (pp. 338–351). IGI Global.

National Academies of Sciences, Engineering, and Medicine. (2018). *How people learn II: Learners, contexts, and cultures*. The National Academies Press. https://doi.org/10.17226/24783

Pekrun, R., Elliot, A. J., & Maier, M. A. (2006). Achievement goals and discrete achievement emotions: A theoretical model and prospective test. *Journal of Educational Psychology*, *98*(3), 583–597. https://doi.org/10.1037/0022-0663.98.3.583

Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *The handbook of research on student engagement* (pp. 259–282). Springer Science and Business Media LLC.

Plass, J. L., Homer, B. D., Mayer, R. E., & Kinzer, C. K. (2020). Theoretical foundations of game-based and playful learning. In J. L. Plass, R. E. Mayer, & B. D. Homer (Eds.), *The Handbook of game-based learning* (pp. 3–24). The MIT Press.

Python Core Team (2015). *Python: A dynamic, open source programming language*. Python Software Foundation. https://www.python.org/

R Core Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ .

Rouet, J., Britt, M., & Potocki, A. (2019). Multiple-text comprehension. In J. Dunlosky & K. Rawson (Eds.), *The Cambridge handbook of cognition and education* (pp. 356–380). Cambridge University Press.

Rowe, J. P., Shores, L. R., Mott, B. W., & Lester, J. C. (2011). Integrating learning, problem solving, and engagement in narrative-centered learning environments. *International Journal of Artificial Intelligence in Education*, *21*(1), 115–133. https://doi.org/10.3233/JAI-2011-019

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, *43*(3), 450–461. https://doi.org/10.1037/0022-3514.43.3.450

Sabourin, J. L., & Lester, J. C. (2014). Affect and engagement in game-based learning environments. *IEEE Transactions on Affective Computing*, *5*(1), 45–56. https://doi.org/10.1109/T-AFFC.2013.27

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In A. Duchowski (Ed.), *Proceedings of the Eye Tracking Research and Applications Symposium 2000* (pp. 71–78). New York, NY: ACM.

Scrimin, S., & Mason, L. (2015). Does mood influence text processing and comprehension? Evidence from an eye-movement study. *British Journal of Educational Psychology*, *85*(3), 387–406. https://doi.org/10.1111/bjep.12080

SMI Experiment Center. (2014). *3.4. 165 [Apparatus and Software].*

Stöckli, S., Schulte-Mecklenbeck, M., Borer, S., & Samson, A. C. (2018). Facial expression analysis with AFFDEX and FACET: A validation study. *Behavior Research Methods*, *50*(4), 1446–1460. https://doi.org/10.3758/s13428-017-0996-1

Storbeck, J., & Clore, G. L. (2005). With sadness comes accuracy; with happiness, false memory: Mood and the false memory effect. *Psychological Science*, *16*(10), 785–791. https://doi.org/10.1111/j.1467-9280.2005.01615.x

Sweller, J. (1988). Cognitive Load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

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. https://doi.org/10.1016/j.learninstruc.2017.08.005

Taub, M., Azevedo, R., Rajendran, R., Cloude, E. B., Biswas, G., & Price, M. J. (2021). How are students' emotions related to the accuracy of cognitive and metacognitive processes during learning with an intelligent tutoring system? *Learning and Instruction*, *72*, 101200. https://doi.org/10.1016/j.learninstruc.2019.04.001

Taub, M., Mudrick, N., Bradbury, A. E., & Azevedo, R. (2020). Self-regulation, self-explanation, and reflection in game-based learning. In J. Plass, B. Horner, & R. Mayer (Eds.), *Handbook of game-based learning* (pp. 239–262). MIT Press.

Taub, M., Sawyer, R., Lester, J., & Azevedo, R. (2019). The impact of contextualized emotions on self-regulated learning and scientific reasoning during learning with a game-based learning environment. *International Journal of Artificial Intelligence in Education, 30,* 97–120. https://doi.org/10.1007/s40593-019-00191-1

Trevors, G., & Kendeou, P. (2020). The effects of positive and negative emotional text content on knowledge revision. *Quarterly Journal of Experimental Psychology*, *73*(9), 1326–1339. https://doi.org/10.1177/1747021820913816

Trevors, G. J., Kendeou, P. A., & Butterfuss, R. (2017). Emotion processes in knowledge revision. *Educational Psychology*, *54*(5–6), 406–426. https://doi.org/10.1080/0163853X.2017.1312201

Trevors, G. J., Muis, K. R., Pekrun, R., Sinatra, G. M., & Winne, P. H. (2016). Identity and epistemic emotions during knowledge revision: A potential account for the backfire effect. *Discourse Processes*, *53*(5–6), 339–370. https://doi.org/10.1080/0163853X.2015.1136507

van den Broek, P., & Helder, A. (2017). Cognitive processes in discourse comprehension: Passive processes, reader-initiated processes, and evolving mental representations. *Discourse Processes*, *54*(5–6), 360–372. https://doi.org/10.1080/0163853X.2017.1306677

Van Moort, M. L., Koornneef, A., & van den Broek, P. W. (2018). Validation: Knowledge- and text-based monitoring during reading. *Discourse Processes*, *55*(5–6), 480–496. https://doi.org/10.1080/0163853X.2018.1426319

Venables, W. N., & Ripley, B. D. (Eds.). (2002). *Modern applied statistics with S*. Springer.

Zaccoletti, S., Altoè, G., & Mason, L. (2019). The interplay of reading-related emotions and updating in reading comprehension performance. *British Journal of Educational Psychology, 90*(3),663–682. https://doi.org/10.1111/bjep.12324