Achievement Emotions with Location-Based Mobile Augmented Reality: An Examination of Discourse Processes in Simulated Guided Walking Tours

Eric G. Poitras, Jason M. Harley and Yang S. Liu

Eric Poitras is an Assistant Professor of Instructional Design and Educational Technology at the University of Utah. Jason Harley is an Assistant Professor in the Department of Educational Psychology at the University of Alberta. Yang Liu is a post-doctoral fellow in the Department of Educational Psychology at the University of Alberta. Address for correspondence: Eric Poitras, Department of Educational Psychology, 1721 Campus Center Drive, Room 3245, Salt Lake City, UT 84112. Email: eric.poitras@utah.edu

Abstract

The purpose of this study is to experimentally manipulate discourse processes hypothesized to impact the emotions students experience when interacting with handheld augmented reality devices in informal learning settings. Research conducted in the field is often limited by practical constraints, requiring heavy investments in time and resources to collect data from large samples of students. To demonstrate the feasibility of our proposed method, a guided walking tour with 60 students using a location-based augmented reality app was simulated in the context of a controlled laboratory setting. The difference between groups of students clustered into distinct profiles of positive and negative self-reported emotions was attributed to patterns in the mined dialogue between students and tour guide. Furthermore, student engagement predicted the ability to recall topics covered in the tour. We discuss the implications and directions for future research in tour simulations conducted in a laboratory setting as a means to evaluate the role of mobile technologies in enhancing learning and desirable emotions.

Keywords: Emotion, Augmented Reality, Educational Data Mining as an Inquiry Method, Informal Learning, Mobile Learning

Introduction

In recent years, there has been an increasing amount of literature on augmented reality (AR) in education as a means to establish design guidelines and instructional practices that enhance learning (Dunleavy, Dede, & Mitchell, 2009; Klopfer & Squire, 2008; Squire & Jan, 2007; Squire, 2010). According to Klopfer (2008), AR devices can be defined as any technology that overlays digital content with information situated in real world settings. Several studies suggests that location-based AR applications, where learning is facilitated through interactive multimedia content that is sensitive to the context or location of the user, is favorable to positive and meaningful learning (see Dunleavy and Dede, 2014). The affordances of the medium may be attributed to their design, providing opportunities to challenge learners by problematizing the subject matter, gamifying instruction through narrative and role-playing, as well as promoting a sense of curiosity about complex topics under investigation (Dunleavy, 2014). For instance, in the *Acropolis Museum* tour, children are instructed to help a horse character to find his friends and get them back to the past. In the case of their parents, the role-playing scenario enables them to learn from the same exhibits about Athenian society and to view digital reconstructions of objects as they were situated in their past architectural setting (Keil et al., 2013).

Empirical research studies that have employed AR in education often focus on early-stage development, usability, and design considerations of the user interface (Poitras, Harley, Compeau, Kee, & Lajoie, 2016; Wu, Lee, Chang, & Liang, 2013). There is a pressing need for controlled and comprehensive evaluation studies of Mobile AR-based historical tours with larger samples and valid measurements that are grounded in theoretical frameworks of learning and engagement (Harley, Poitras, Jarrell, Duffy, & Lajoie, 2016; Harley, Lajoie, Tressel, & Jarrell, 2018). The issue has grown in importance in light of recent advances in AR, making such technologies more accessible in informal learning environments such as museums and historical heritage sites, where conducting controlled studies is challenging for practical reasons and limited resources (Bressler, 2013; Tallon, 2013). There are three primary aims of this study: (1) To remedy the time and resource intensive nature of field work by evaluating a location-based AR application in the context of a walking guided tour that is simulated in the context of a laboratory; (2) to systematically manipulate the dynamic social interaction that occurs between the tour guide and visitor; and (3) to discover novel patterns via data mining that are indicative of learning and affective engagement. In the following sections, we elaborate further on the notion of simulated tours and demonstrate the use of these methods towards the evaluation of conversational scaffolds aimed at supporting positive learning experiences with the benefit of location-based AR applications. The term conversational scaffold refers to the type of discourse strategy used by an instructor, in this case, a tour guide, to support learners in processing content and experiencing positive emotions.

Research Background

The underlying assumption of this study is that simulated guided walking tours through historical heritage sites and museum exhibits not only support learning, but may also provide unique opportunities for researchers to study how AR applications overlay digital content to physical objects or environments. Prior studies have shown that increased presence, or a sense of the realism of the AR display, under different augmented reality devices is associated to lower cognitive load when learning chemistry (Chen, Wang, & Chiang, 2009). Furthermore, learner motivation and interest typically increases in immersive environments, leading to increased levels of interaction, immersion, and imagination, which are found to promote collaborative learning (Huang, Rauch, & Liaw, 2010) and learning efficiency (Yim & Seong, 2010). A major advantage of virtual immersive experiences is to promote the subjective impression that learning is situated in a comprehensive and realistic experience (Dede, 2009). For example, hands-on and interactive digital exhibits in museums allow learners to interact with virtual characters (Roussou, 2004). Immersive interfaces may also facilitate learning by simulating real-world contexts such as virtual field trips facilitated through geospatial technologies such as Google Earth (Trautmann & MaKinster, 2010). Learners may walk through a scene to see photographs from multiple angles while zooming in and out of each of them. In this study, we apply web-based virtual tours with interactive 180 degree 3D panoramic photographs of locations to investigate the impact of emotions on learning with AR.

The study of emotions is a critical (but under-explored) area of research within the field of AR in education due to the immersive, motivating, engaging affective qualities of the medium (Akcayir & Akcayir, 2017; Bacca, Baldiris, Fabregat, Graf, and Kinshuk, 2014; Diegmann, Schmnidt-Kraepelin, Eynden, & Basten, 2015) and the relationship between learning and emotions (Pekrun & Perry, 2014). Activity-related achievement emotions are experienced as a result of attention being directed toward an achievement task itself (e.g., a particular course being boring) as opposed to the outcome of a task (e.g., pride from performing well on a test; Pekrun, 2006). According to control-value theory of achievement emotions (Pekrun, 2006; Pekrun and Perry, 2014), learners' emotional states fluctuate on the basis of subjective appraisals of control and value. For example, learners who feel in control of and value an activity are more likely to feel positive emotions, such as enjoyment. In contrast, learners who do not value an activity are more likely to experience boredom, regardless of their own perception of autonomy and whether they can make choices in the context of a guided walking tour. Previous research has shown that positive activating emotions (i.e., higher physiological arousal), such as enjoyment, are typically related to adaptive learning outcomes rather negative deactivating emotions such as boredom (Pekrun and Perry, 2014). These relationships may partly be explained by the mediating role of attentional resources allocated to processing the subject matter taught in the tour environment. Learners who experience a higher degree of enjoyment may allocate more attention to the task, enabling deep processing of the subject matter (Pekrun, Frenzel, Goetz, & Perry, 2007; Pekrun, Goetz, Daniels, Stupnisky, & Raymond, 2010; Mann & Robinson, 2009; Watt & Vodanovich, 1999).

An important implication of control-value theory is that guided walking tours enhanced by AR applications can be designed to promote learners' perception of control and value appraisals via conversational scaffolds, thereby leading to more positive emotions and adaptive learning behaviors. Much uncertainty still exists about the relationship between the tour guide and learners, and how this may impact learner appraisals. The design of AR enhanced tours often lack manipulation of relevant characteristics under controlled conditions, such as how learners navigate through locations, the role-playing and narrative that unfolds throughout learning, as well as the task objective and conditions (e.g., time on task, materials, learner community, Poitras et al., 2016). In this study, we manipulate conversational scaffolds used by tour guides in the context of the simulated walking tour. In doing so, we rely on an intelligent quantitative field observation system that serves as a research tool called mObserver. The mObserver allows researchers playing the role of a tour guide to systematically manipulate conversational scaffolds while recording behavioral observations, namely, the discourse moves used by the guide to enhance learning. Individual moves are randomly assigned to each learner by the tour guide as displayed in the system user interface upon visiting a particular tour location.

Research Objectives and Questions

This study follows the aforementioned method to simulate guided walking tours with location-based AR as a means to address the following research questions: (RQ1) Are there distinct activity-related emotion profiles that characterize visitors' experiences in a simulated tour?; (RQ2) Do tour guide's discourse moves explain visitors' belonging to different activity-related emotion profiles?; and (RQ3) Is there a relationship between activity-related emotion profiles as well as visitors' ability to recall and understand topics covered during the tour? We systematically manipulate the conversational moves used by the tour guide, based on previous examination of discourse processes in similar settings (Harley et al., 2016), with the expectation that at least two distinct profiles would emerge from the self-report data based on subjective appraisals of goal attainment (Jarrell, Harley, Lajoie, & Naismith, 2017; Carver & Sheier, 2014; D'Mello et al., 2014), referring to either positive or negatively-valenced emotions. We also hypothesized the possibility of a third, low emotional intensity profile, in-line with previous research clustering learners' emotions from interacting with technology-rich learning environments (Jarrell et al., 2017). Our second research hypothesis states that there should be meaningful and consistent patterns between the tour guide's discourse moves and visitor's emotion profiles. The second research question this hypothesis addresses was largely exploratory on account of the novelty of the proposed methodology (including the large number of potential combinations of features) and lack of theoretical guidance. Finally, we expect that students who experience positive emotions are more likely to outperform those who report negative emotions across learning outcome measures (Pekrun & Perry, 2014).

Methods

Participants

The sample included 60 students (Female = 39; Male = 21) enrolled in the research participation pool of the Department of Educational Psychology at a public university in the western region of the United States. The sample had a mean age of 24 years old (*SD* = 5) a GPA of 3.3 (*SD* = 0.58) and had completed an average of two years in their program of study (*M* = 2.2; *SD* = 1.2).

Research Design

A single-group design was followed in this study to examine the impact of scaffolding prompts delivered by a tour guide on students' understanding of the subject matter and achievement emotions (see Table 1). This study focuses on the log trace data obtained from the mObserver app as well as the self-report measures of students' topic understanding and retrospective activity-based achievement emotions. The mObserver app selected randomized sequences of prompts and displayed these prompts to the tour guide on the basis of the tour location under examination. To transition from one location to the next, a total of five prompts (i.e., transitional statements) were delivered by the tour guide to each student at the end of each location visit. Once a tour location was visited by a student, the tour guide had the opportunity to scaffold students' understanding of the subject matter through a subset of 19 random prompts proposed by the mObserver app (i.e., open-ended questions and instructions).

Table 1

Sample of Scaffolding Prompts Delivered by the Tour Guide

Prompts
What do you know about the life of Dr. Milton Bennion?
Could you summarize what you currently know about Dr. Milton Bennion?
Do you feel that you understand the documents and are ready to move on?
Can you guess what was the purpose of the outdoor area shown in the interactive video?
Could you summarize what you currently know about the dedication of the building to Dr. Milton Bennion?
Do you feel that you understand the documents and are ready to move on?

You can move this viewport closer to see the portrait of Dr. Milton Bennion.
Pay close attention to the excerpts taken from documents written by Dr. Milton Bennion, I would recommend to read them over a few times.
Can you say in your own words what was Dr. Milton Bennion saying in these documents.

Location-Based AR Mobile App

The DiscoverUofU app for mobile and wearable devices (i.e., Android, iOS, Microsoft, Glass) is an interactive audio guide for historic figures and landmarks surrounding a university campus. The app delivers multimedia content (i.e., pictures, texts, audio) on the basis of GPS coordinates that correspond to each tour location. This study examined the tour entitled the "Life of Milton Bennion", which features an audio narration drawing upon excerpts of historical multimedia and landmarks that pertain to Dr. Milton Bennion's beliefs, values, and achievements in the mid-1900s. The tour consists of four locations visited in the following order: life and dedication (i.e., South East); Milton Bennion Hall (i.e., West area); Dr. Milton Bennion's values (i.e., North East); and Dr. Milton Bennion's values (i.e., Indoors area). The materials were adapted from the multimedia archives of the J. Willard Marriott Library at the University of Utah as well as Campbell (1990).

Measures

Topic Understanding. Following Royer, Carlo, Dufresne, and Mestre (1996) as well as Wiley and Voss (1999), a sentence and inference verification task was created to assess students' understanding of the historical subject matter (see Table 2 and 3). The sentence verification task consisted of 16 binary-response items (i.e., eight true and eight false response items) that assessed students' understanding of the meaning of textual segments. Students were instructed to indicate whether the idea conveyed in a revised textual segment (i.e., original, paraphrase, meaning change, and distractor) was similar to the information they read about Dr. Milton Bennion. The inference verification task consisted of eight binary-response items, which students indicated whether the statement (i.e., near and far inferences that are either valid or not) was true on the basis of the information they read.

Table 2

Sample of Response Items for Sentence Verification Task (SVT)

Item	Original	Category	Location
Milton Bennion was a member of the University of Utah faculty for 40 years, including 20 years as dean of the School of Education and a year as vice-president.	Milton Bennion was a member of the University of Utah faculty for 40 years, including 20 years as dean of the School of Education and a year as vice-president.	Original	South East
Milton Bennion was the dean of the School of Education from 1913 to 1941, while also serving as vice president at the request of Dr. George Thomas in 1940 and again in 1941.	From 1913 to 1941, he was dean of the School of Education and also was vice president under Dr. George Thomas in 1940 and 1941.	Paraphrase	South East
Milton Bennion joined the University of Utah faculty in 1901 as assistant professor of education and was named to an associate professorship in 1904.	He joined the University of Utah faculty in 1901 as assistant professor of education and was named to a full professorship in 1904.	Meaning Change	South East
The Faculty of the University of Utah learned several years after of the death of Dr. Milton Bennion, which occured April 5th, 1953, in Salt Lake City.	The Faculty of the University of Utah were deeply grieved to learn of the death of Dr. Milton Bennion, which occured April 5th, 1953, in Salt Lake City.	Distractor	South East

Table 3

Sample of Response Items for Inference Verification Task (IVT)

Item	Original	Category	Location

Milton Bennion was a senior faculty member with an exemplary record of service to the University of Utah.	All text excerpts	Near & Valid Inference	South East
Milton Bennion and his family immigrated to Utah at the turn of the 20th century.	All text excerpts	Near & Invalid Inference	South East
Milton Bennion was awarded an honorary doctoral degree for his contributions to the University of Utah.	All text excerpts	Far & Valid Inference	South East
Milton Bennion retired from the University of Utah and served during the second World War.	All text excerpts	Far & Invalid Inference	South East

Achievement Emotions. A self-report measure assessed learners' levels of three emotions: enjoyment, anger, and boredom. After the learning session participants answered a 36-item version of the measure, which asked them about their level of enjoyment, boredom, and anger toward each component (Guide, Tour, Learning; see Table 4 in the Results section) and tour locations. The questionnaire (Harley et al., 2016, 2018) was an adapted version of the Academic Achievement Emotion Questionnaire (Pekrun et al., 2002; 2011) and used a five-point likert scale. We selected these three emotional states based on the prevalence of emotions in technology-rich learning environments and theoretical considerations (D'Mello et al., 2013; Pekrun & Perry, 2014). We chose to measure only three emotions to avoid item fatigue on account of the frequency with which learners were asked to report their emotions, and doing so, required them to report these discrete emotions toward three different components of their learning session. Moreover, research has found low incidents of negative activating emotions in interactions with mobile AR apps, similar to the one this study focused upon (Harley et al., 2016, 2018). While our resulting three emotions intentionally occupy three of the four quadrants of an affective circumplex model (valence by activation; Russell, Weiss, & Mendelsohn, 1989), measuring discrete emotions permits generalizing findings to both these specific states as well as their higher-level quadrants (Harley, Bouchet, Hussain, Azevedo, & Calvo, 2015; Pekrun, Lichtenfeld, Marsh, Murayama, & Goetz, 2017), is still aligned with the dominant theory of emotions in education, the control-value theory of achievement emotions (Pekrun, 2006; Pekrun & Perry, 2014), and allowed us to use an instrument based on the AEQ and previously examined in studies with similar technology. Making theoretical assumptions, measurement methods, and data analyses decisions explicit is important for minimizing potential confusion in the emotion literature (Weidman, Steckler, & Tracy, 2017), particularly, when dealing with the treatment of dimensional and discrete emotions.

Table 4

Sample of Items for Retrospective, Activity-related Achievement Emotions Questionnaire (AEQ)

Item	Category	Location
After visiting this location, I started to look forward to the next one.	Enjoyment	South East
Because I got bored at this location, my mind began to wander.	Boredom	South East
I felt angry while visiting that particular location.	Anger	South East

Experimental Procedure

The tour was simulated in a laboratory setting. Students completed a consent form and demographic questionnaire prior to their participation in the study. They were instructed on the use of the DiscoverUofU app as well as the three-dimensional and animated panoramic pictures of each tour location. The students then completed the tour by visiting each tour location in a serial manner. After visiting a location and prior to moving on to the next, the students answered questions asked by the tour guide. At the end of the visit, students completed the self-report measure of achievement emotions that referred to to each tour location, the sentence and inference verification tasks, as well as a brief argumentative essay.

Results

We organized our findings according to the research questions outlined earlier in this paper: (RQ1) discovering activity-related emotion profiles; (RQ2) explaining profile membership on the basis of the conversational scaffolds used by the tour guide; and (3) examining the impact of activity-related emotion profiles towards topic recall and understanding (RQ3).

RQ1: Profiles of Activity-Related Emotions in Simulated Guided Walking Tours

Figure 1 shows the silhouette plots of clusters obtained from the k-Means algorithm for students' activity-related emotions (anger, enjoyment, boredom) towards the tour, guide, and location. The k-Means algorithm was chosen due to the continuous scale of the variables as well as the Euclidean distance as the similarity measure. A higher score indicates a better solution as determined by the distance of each data point to the centroid, ranging on a scale of -1 to 1 (see Rousseeuw, 1987). A two-cluster solution obtained the highest silhouette score of 0.3 and was retained for the purposes of mining discourse patterns to discover subgroups. The two-cluster solution was further validated through the best fitting model obtained from 1000 separate instances of k-Means using randomly assigned centroids and specifying an f(k) criterion of 0.85 (Pham, Dimov, and Nguyen, 2004; see Figure 2). Figure 3 shows the boxplot of activity-related emotions across the clusters, labeled as positive and negative emotion profiles. The examination of the mean and 95% confidence intervals confirm that there is a statistically significant difference between profiles obtained on all activity-related emotions, where the positive emotions profile is characterized by a higher degree of enjoyment and a lower degree of boredom and anger directed toward the tour guide, and location.

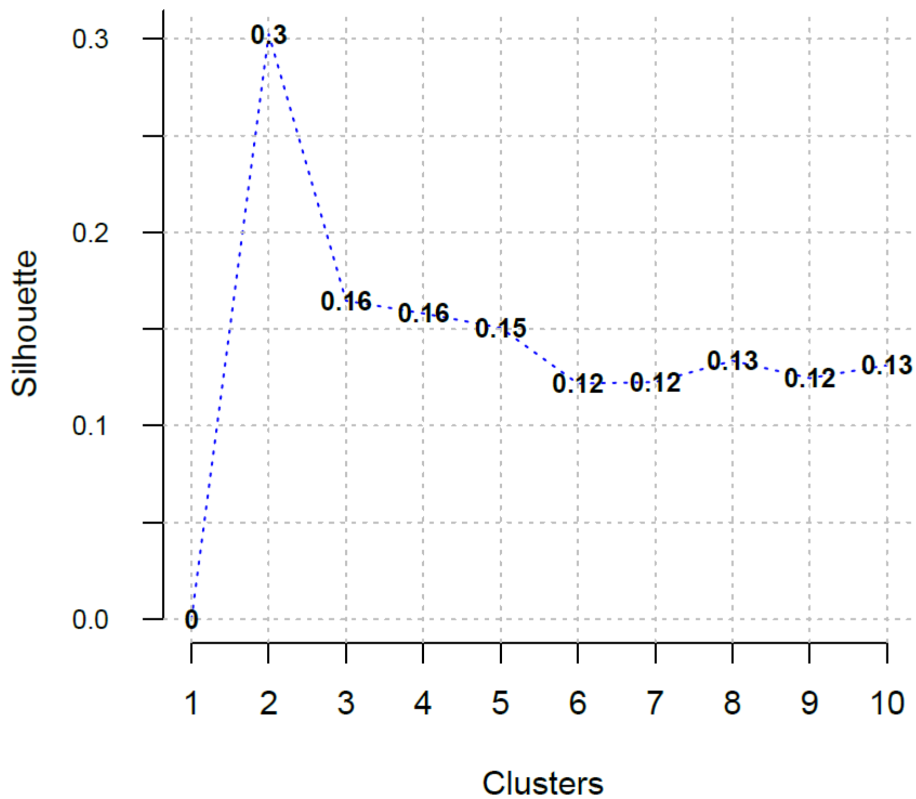

Figure 1. Silhouette plot of the K-means clustering algorithm for activity-related emotions (anger, boredom, enjoyment) towards the tour, guide, and location. The 2-cluster solution obtained the highest score (0.3) on the scale ranging from -1 to 1, and was thus used for further analyses given that the solution minimized the distance between the centroids and data points.

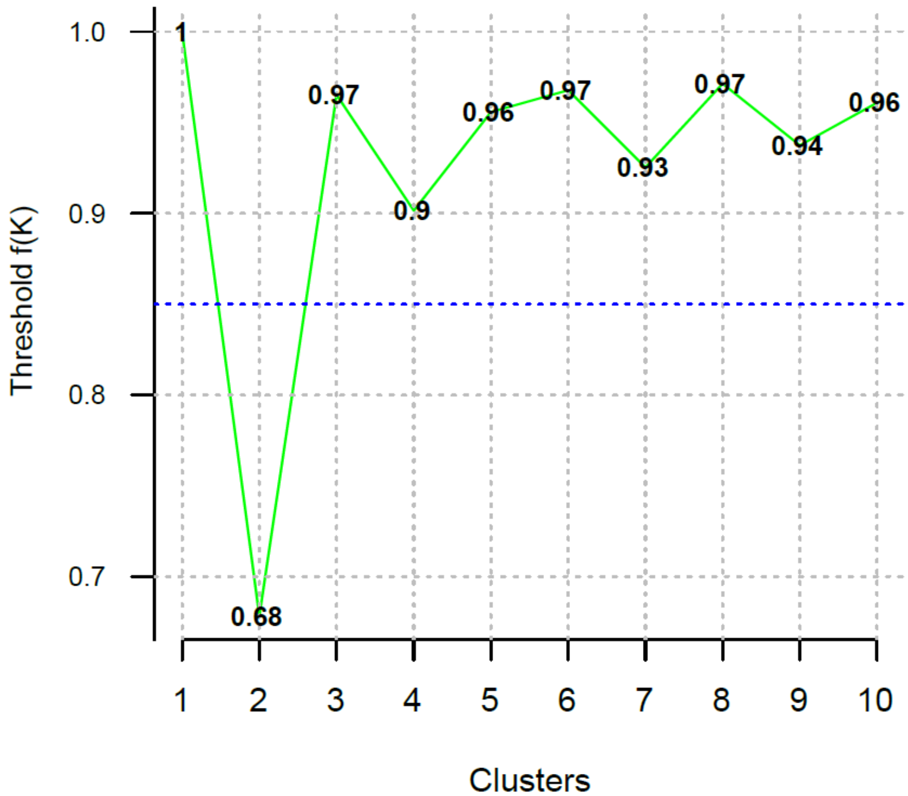

Figure 2. Best fitting model of 1000 separate instances of k-means using f(k) criterion. The 2-cluster solution obtained the lowest score (0.68) and was retained for further analysis.

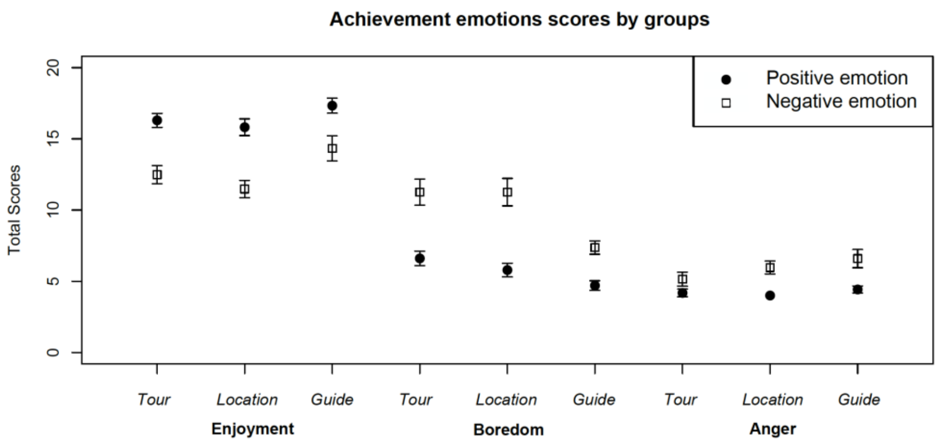

Figure 3. Activity-related emotion scores across positive and negative emotion profiles. The confidence intervals on all measures indicate significant differences across all scales of activity-related emotions.

We now turn our attention to the examination of discourse processes as a means to explain why visitors were more or less likely to experience positive and negative emotions during the AR enhanced tour.

RQ2: Subgroups of Discourse Processes in the Positive and Negative Emotion Profiles

Subgroup discovery algorithms (Wrobel, 1997; Klösgen, 2002) aim to induces rules that are both generalizable and of interest to a set of learning behaviors and a phenomena under investigation. In this study, a rule consists of a premise (i.e., discourse move(s)) that are able to account for membership in a particular emotion profile (i.e., either positive or negative). The algorithm followed an exhaustive search process, examining combinations of up to 6 discourse moves as premises in the generated rules. Furthermore, the rules induced from the analysis had to cover at least 20% of the examples in the dataset and were optimized in terms of predictive accuracy.

Table 5 shows the rules induced from the subgroup mining algorithm. Examining the most predictive rule, this discourse pattern observed in the AR enhanced tour was associated to visitors' positive emotions. A total of 11 visitors experienced positive emotions, while 2 of them reported negative emotions, accounting for 22% of the total amount of visits. As such, this particular rule predicted emotion profile membership with 85% precision and 70% accuracy. In order to interpret the rule premises, the corresponding discourse moves are listed below:

4. During what time period did Dr. Milton Bennion work at the University of Utah?

6. Do you feel that you understand the documents and are ready to move on?

16. Can you summarize what you now know regarding Dr. Milton Bennion's beliefs?

In asking these questions to the visitors, the tour guide was supporting them to engage in both cognitive and metacognitive processes. The tour guide prompted the visitors to recall what they had previously learned about Dr. Milton Bennion, establishing a timeline of events that unfolded through the narrative shown in the AR app. The visitors were also asked to draw inferences from the information provided in the documents, elaborating further on the beliefs of Dr. Milton Bennion based on an excerpt of a transcript of a speech. Finally, visitors were prompted to engage in metacognitive monitoring activities by judging their own learning of the subject matter covered at a particular location. The 13 visitors who were asked these questions by the tour guide tended to experience positive emotions directed towards the tour.

Table 5

Subgroups of discourse processes based on student emotion profiles

Premise	Conclusion	Pos	Neg	Size	Cov	Pre	Acc	Length
4 = true, 6 = true, 16 = true	Positive	11	2	13	0.22	0.85	0.70	3
6 = true, 10 = true, 16 = true	Positive	11	2	13	0.22	0.85	0.70	3
6 = true, 10 = true	Positive	12	5	17	0.28	0.71	0.67	2
6 = true, 16 = true	Positive	12	5	17	0.28	0.71	0.67	2
14 = true, 22 = false	Positive	12	5	17	0.28	0.71	0.67	2

Note: The premise of each rule defines the number or ID of the discourse move used by the tour guide. The premises are used to determine whether one move is used or not to predict whether the emotion profile a student belongs to is either positive or negative (i.e., the classification labeled as the "Conclusion" column). The following notation is used to refer to the metrics to assess predictive accuracy: Pos (Positive instances covered by the rule); Neg (Negative instances not covered by the rule); Size (Total amount of instances covered by the rule); Cov (Coverage or percentage of instances covered by the rule); Pre (Precision or Positive instances given the size of instances covered by the rule); Acc (Accuracy or positive and negative instances covered by the rule given those also not covered by the rule).

The use of subgroup discovery mining in conjunction with the randomized delivery of discourse moves by tour guides thus allows for finer grained investigations of processes that are predictive of either positive or negative emotions in AR enhanced tours. The following section examines the impact of visitors' emotional experiences towards their own learning.

RQ3: Impact of Emotion Profiles towards Content Knowledge

Table 6 shows the Welch t-tests of student learning outcome scales and sub-scales between activity-related emotion profiles (positive versus negative). There was a statistically significant main effect of emotion profile on students' ability to correctly identify distractor statements where students in the positive emotion profile (M = 2.39) outperformed those in the negative emotion profile (M = 1.85), $t(58) = 2.15$, $p < .05$.

Table 6

Welch t-tests of student learning outcomes across activity-related emotion profiles

Variables	Emotion Profiles				
	Positive	Negative	*t*	*df*	*p*
Sentence Verification Task (/16)	9.97	9.81	0.37	52.4	0.71
1.Original (/4)	3.12	3.44	-1.81	49.3	0.08
2.Paraphrase (/4)	3.42	3.33	0.49	48.7	0.63
3.Meaning Change (/4)	1.03	1.19	-0.70	53.6	0.48
4.Distractor (/4)	2.39	1.85	2.15	58.0	0.04*
Inference Verification Task (/8)	5.61	5.41	0.57	50.4	0.57
1.Near Valid (/2)	1.55	1.44	0.68	55.1	0.50
2.Near Invalid (/2)	1.73	1.63	0.73	49.4	0.47
3.Far Valid (/2)	1.00	0.81	0.93	54.6	0.36
4.Far Invalid (/2)	1.33	1.52	-1.15	53.8	0.26

This finding provides some preliminary support that students' activity-related emotions are related to their understanding of the historical topic under examination using mobile AR apps, which may be determined by the type of conversational scaffold used by the tour guide. We discuss the merits and limitations of the proposed approach and the broader implications of these findings in the following section.

Discussion

This study set out to investigate whether activity-related emotion profiles could be identified in the context of an Mobile AR-based guided walking tour simulated in the laboratory (RQ1), and whether these profiles could be accounted for by the conversational scaffolds of a tour guide (RQ2) and would impact topic recall and understanding (RQ3). We found that two distinct profiles of activity-related emotions characterized students' experience with the location-based AR application in the context of the guided walking tour simulated in the laboratory setting. The observed difference in activity-related emotions was characterized as being due to their valence, where participants reported higher levels of positive emotions across tour locations than negative ones. A larger number of clusters, such as a low intensity one or a cluster pattern differentiating boredom and anger may have emerged with higher participant ratings of these negative emotions. While a two-cluster, valence-based solution provides less granular insight than one that might have better reflected the nuances between the negative emotions evaluated, our findings align with emotion theory and replicate findings from other studies concerning the relationship between valence clustering and performance with a novel type of mobile AR environment. Namely, one where conversational scaffolds used by the tour guide are manipulated for the purposes of appraising their specific contributions towards observed knowledge of the content gained throughout the guided walking tour in a simulated setting (Harley et al., 2016).

The activity-related emotion profiles were also associated to patterns in the discourse moves used by the tour guide, as hypothesized. This is important because such associations may be used to predict learners' emotions and, accordingly, help inform which messages are delivered based on a more informed student model of the learners' hypothesized state. The observed associations between specific conversational scaffolds used by the tour guide and activity-related emotions might be explained in terms of the appraisals made by the participants towards the tour experience. It may be that certain learner-system discussions pertaining to tour topics lead to, for example, higher appraisals of intrinsic value of the tour or heightened experiences of autonomy (i.e., control) related to how the material was presented to them.

Our third hypothesis was partially supported where the only significant relationship to emerge was in the expected direction: learners classified as belonging to the positive emotion cluster outperformed those in the negative emotion cluster in identifying distractor statements. It is possible that the relatively low mean levels of negative emotions reported by participants were not intense enough to hinder performance, but that the higher levels of positive emotions may have been sufficient to help foster learning. Furthermore, it may also be that identifying distractor statements is particularly cognitively demanding relative to the other learning outcomes and therefore performance on this assessment may have been more sensitive to affective tendencies than others.

This combination of findings provides some support for the conceptual premise that simulated guided walking tours with immersive technology are a feasible method to evaluate location-based AR applications. The findings are in agreement with Harley et al. (2016) which showed that achievement emotions are associated with learning outcomes at comparable levels to both simulated and actual tour settings. In the current study, a systematic manipulation of conversational scaffolds used by the tour guide through the use of the intelligent quantitative field observation system provided a more nuanced evaluation.

A limitation of this study is that in the absence of a control condition, as in Harley et al. (2016), means that some caution should be applied to interpreting the findings, as they might not be transferable to an actual tour setting. Further confirmation is required with a control condition where no prompts are delivered by the tour guide to ascertain the differential impacts on knowledge gain. Additionally, the study lacked a formal test of prior knowledge about instructional content which limits assessments of learning, albeit with the caveat that the historical topic taught in the mobile AR app is not commonly known (i.e., not reported widely in local newspapers, and occurred in the 1950s), and is also excluded from classroom instruction. That said, the most important finding to emerge from this study is that this method allows one to reveal fine-grained associations, namely, the degree to which certain discourse moves are associated to distinct emotion profiles. This not only allows for evidence-based improvements to the tour experience, it also allows future research to optimize the tour experience in real-time by implementing algorithms in the system to manipulate, track, and evaluate conversational scaffolds in a principled and strategic manner.

Globally, we have much to learn about the guidelines that should be followed to ensure that simulated tours are sufficiently immersive to emulate real-life conditions. As Poitras et al. (2016) have noted, simulated tours may impact the emotional experiences of learners in a different manner than those in the actual setting. Furthermore, the examination of the process data collected from measures used in the laboratory setting, for example, the verbal protocol, physiological, and biomechanical data (i.e., audio recording of utterances, heart rate monitor, galvanic skin response, and depth camera to track movements and posture combined with a camera to record facial expressions) may reveal additional information to support our claims. It may be the case therefore that fluctuations in process data provide a more nuanced assessment of activity-related emotions as well as changes in them than those obtained from the self-report measures (Harley, 2015). Capturing dynamic process data would thus enable researchers to uncover the most influential conversational scaffolds by tracking their onset in the context of the mObserver application.

Our future work will aim to manipulate the conversational scaffolds in the context of guided walking tours augmented with digital content, but designed in accordance with guidelines for reducing cognitive load from the cognitive theory of multimedia learning (Mayer & Moreno, 2003). Prior research suggests that this approach may result in increased student performance on

AR lessons over control conditions (Yim & Seong, 2010). Furthermore, design guidelines to reduce cognitive load in AR-based interfaces have been shown to reduce divided attention related issues in a range of different tasks, including but not limited to driving with AR windshield displays (Kim & Dey, 2009), medical procedures (Klatzky et al., 2008), and assembly tasks (Tang et al., 2003).

In further research, there is a need to pursue the development of mObserver to enable the evaluation of increasingly sophisticated conversational scaffolds that are grounded in theoretical frameworks of learning (Harley, Lajoie, Frasson, Hall, & 2017) in the context of simulated guided walking tours with location-based AR applications. In these situations, stakeholders are involved in the design process, where the user interface is improved on an iterative basis by incorporating feedback. The quantitative field observation system outlined in this paper stands to inform this approach by providing summaries of learning sessions to support system designers. One promising direction is to allow for branching scenarios in the conversation between tour guides and learners to evaluate not only the most suitable topics to discuss, but also how to discuss them. As we also alluded to in this section, the delivery method of the conversational scaffold should also be subject to further examination. The randomization method used in the current study requires increasingly large sample sizes to ensure adequate coverage of candidate discourse moves. A promising direction may be to investigate optimization methods rather than rely on experimental manipulation to more efficiently search through the space of candidate moves that could be delivered and find the most suitable combinations. Therefore, we call on researchers to consider these methodological concerns in their implementation of simulated tours to evaluate and establish best practices in the design guidelines of location-based AR applications.

Commented [JMH1]: This was already in the reference section; thought it best to put here.

Acknowledgements

The research presented in this paper has been supported by funding from the Social Sciences and Humanities Research Council of Canada. .

Statements on open data, ethics, and conflicts of interest

We declare that we do not have any conflicts of interest regarding the study. An IRB approval was obtained prior to the conduct of the study and informed consent was obtained from participants. Furthermore, any request to access the data should be directed to the first author (Eric Poitras; eric.poitras@utah.edu).

References

Akcayir, M., Akcayir, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educational Research Review, 20,* 1-11.

Bacca, J., Baldiris, S., Fabregat, R., Graf, S., & Kinshuk. (2014). Augmented Reality Trends in Education: A Systematic Review of Research and Applications. *Educational Technology & Society, 17*(4), 133–149.

Bressler, D. M. (2013). Gateways to mobile learning. In Z. L. Berge & L. Muilenburg (Eds.), *Handbook of mobile learning* (pp. 224–234). New York, NY: Routledge.

Campbell, R. F. (1990). Nine lives, leadership and the University of Utah's School of Education, 1869-1988. Graduate School of Education University of Utah. Salt Lake City.

Chen, Y. C., Wang, S. J., & Chiang, Y. L. (2009). Exploring the Effect of Presence in an AR-based Learning Environment. *Proceedings of the 13th Global Chinese Conference on Computers in Education*, Taipei, Taiwan, 693-702.

Dede, C. (2009). Immersive interfaces for engagement and learning. *Science, 323*(5910), 66-69.

Diegmann, P., Schmidt-Kraepelin, M., Eynden, S., Basten, D. (2015). Benefits of Augmented Reality in Educational Environments - A Systematic Literature Review. *Wirtschaftsinformatik Proceedings 2015*. 103.

Dunleavy, M. (2014). Design principles for augmented reality learning. *TechTrends, 58*(1), 28–34. doi:10.1007/s11528-013-0717-2

Dunleavy, M., & Dede, C. (2014). Augmented reality teaching and learning. In J. M. Spector et al. (Eds.), *Handbook of research on educational communications and technology* (pp. 735–745). New York, NY: Springer. doi:10.1007/978-1-4614-3185-5_59

Dunleavy, M., Dede, C., & Mitchell, R. (2009). Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. *Journal of Science Education and Technology, 18*(1), 7–22. doi:10.1007/s10956-008-9119-1

Harley, J. M. (2015). Measuring emotions: A survey of cutting-edge methodologies used in computer-based learning environment research. In S. Tettegah & M. Gartmeier (Eds.). *Emotions, Technology, Design, and Learning* (pp. 89-114). London, UK: Academic Press, Elsevier.Harley, J. M., Bouchet, F., Hussain, S., Azevedo, R., & Calvo, R. (2015). A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. *Computers in Human Behavior, 48, 615-625*.

Harley, J. M., Lajoie, S. P., Frasson, C., & Hall, N. C. (2017). Developing emotion-aware, advanced learning technologies: A taxonomy of approaches and features. *International Journal of Artificial Intelligence in Education, 27*(2), 268-297.

Harley, J. M., Lajoie, S. P., Tressel, T., & Jarrell, A. (2018). Fostering positive emotions and history knowledge with location-based augmented reality and tour-guide prompts. *Learning and Instruction*.

Harley, J. M., Poitras, E. G., Jarrell, A., Duffy, M. C., & Lajoie, S. P. (2016). Comparing virtual and location-based augmented reality mobile learning: Emotions and learning outcomes. *Educational Technology Research & Development, 64*(3), 359-388.

Huang, H. M., Rauch, U., & Liaw, S. S. (2010). Investigating learners' attitudes toward virtual reality learning environments: Based on a constructivist approach. *Computers & Education, 55*(3), 1171-1182.

Jarrell, A., Harley, J.M., Lajoie, S.P., & Naismith, L. (2017). Success, failure and emotions: Examining the relationship between performance feedback and emotions in diagnostic reasoning. *Educational Technology Research and Development, 65(5),* 1263–1284. DOI: 10.1007/s11423-017-9521-6

Keil, J., Pujol, L., Roussou, M., Engelke, T., Schmitt, M., Bockholt, U., & Eleftheratou, S. (2013). A digital look at physical museum exhibits: Designing personalized stories with handheld Augmented Reality in museums. In *Digital heritage international congress (DigitalHeritage), 2013* (Vol. 2, pp. 685–688). Marseille, France: IEEE; doi:10.1109/DigitalHeritage.2013.6744836

Kim, S. J., & Dey, A. K. (2009). Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI) 2009*. 133-142.

Klatzky, R. L., Wu, B., Shelton, D., & Stetten, G. (2008). Effectiveness of augmented-reality visualization versus cognitive mediation for learning actions in near space. *ACM Transactions on Applied Perception (TAP), 5*(1), 1-23.

Klopfer, E. (2008). *Augmented learning: Research and design of mobile educational games.* Cambrigde, MA: MIT Press. doi:10.7551/mitpress/9780262113151.001.0001

Klopfer, E., & Squire, K. (2008). Environmental Detectives – the development of an augmented reality platform for environmental simulations. *Educational Technology Research and Development, 56*(2), 203-228.

Klösgen, W. (2002). Subgroup discovery. In W. Klösgen and J. Zytkow (Eds.), *Handbook of data mining and knowledge discovery*. New York: Oxford University Press.

Mann, S., & Robinson, A. (2009). Boredom in the lecture theatre: An investigation into the contributors, moderators, and outcomes of boredom amongst university students. *British Educational Research Journal, 35*(2), 243-258.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43-52.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18,* 315-341.

Pekrun, R., Frenzel, A. C., Goetz, T., Perry, R. P. (2007). The control-value theory of achievement emotions : An integrative approach to emotions in education. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in education* (pp. 13-36). San Diego, CA : Academic Press.

Pekrun, R., Goetz, T., Daniels, L., Stupnisky, R. H., & Raymond, P. (2010). Boredom in achievement settings : Exploring control-value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology, 98*(3), 583-597.

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary educational psychology*, *36*(1), 36-48.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational psychologist*, *37*(2), 91-105.

Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement emotions and academic performance: Longitudinal models of reciprocal effects. *Child Development*, *88*(5), 1653-1670.

Pekrun, R., Perry, R. P. (2014). Control-value theory of achievement emotions. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International Handbook of Emotions in Education* (pp. 120-141). New York, NY: Taylor & Francis.

Pham, D. T., Dimov, S. S., & Nguyen, C. D. (2004). An incremental K-means algorithm. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, *218*(7), 783-795.

Poitras, E. G., Harley, J. M., Compeau, T., Kee, K., & Lajoie, S. P. (2016). Augmented reality in informal learning settings: Leveraging technology for the love of history. In R. Zheng & M. K. Gardner (Eds.), *Handbook of Research on Serious Games for Educational Applications*. IGI Global.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, *20*, 53-65.

Roussou, M. (2004). Learning by doing and learning through play: An exploration of interactivity in virtual environments for children. *ACM Computers in Entertainment*, *2*(1), 1-23.

Royer, J. M., Carlo, M. S., Dufresne, R., & Mestre, J. (1996). The assessment of levels of domain expertise while reading. *Cognition and instruction*, *14*(3), 373-408.

Squire, K. (2010). From information to experience: Place-based augmented reality games as a model for learning in a globally networked society. *Teachers College Record, 112*(10), 2565–2602.

Squire, K., & Jan, M. (2007). Mad city mystery: Developing scientific argumentation skills with a placebased augmented reality game on handheld computers. *Journal of Science Education and Technology, 16*(1), 5–29. doi:10.1007/s10956-006-9037-z

Tallon, L. (2013). *Mobile strategy in 2013: An analysis of the annual museums & mobile survey*. Retrieved from http://www.museums-mobile.org/survey

Tang, A., Owen, C., Biocca, F., & Mou, W. (2003). Comparative effectiveness of augmented reality in object assembly. In V. Bellotti, T. Erickson, G. Cockton, & P. Korhonen (Eds.), *Conference on Human Factors in Computing Systems - Proceedings* (pp. 73-80).

Trautmann, N., & MaKinster, J. (2010). Flexibly adaptive professional development in support of teaching science with geospatial technology. *Journal of Science Teacher Education, 21*(3), 351–370.

Watt, J. D., & Vodanovich, S. J. (1999). Boredom proneness and psychosocial development. *The Journal of Psychology, 133*, 303-314.

Weidman, A. C., Steckler, C. M., & Tracy, J. L. (2017). The jingle and jangle of emotion assessment: Imprecise measurement, casual scale usage, and conceptual fuzziness in emotion research. *Emotion, 17*(2), 267.

Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology, 91*(2), 301.

Wrobel, S. (1997). An algorithm for multi-relational discovery of subgroups. In *Proceedings of the first European symposium on principles of data mining and knowledge discovery* (pp. 78-87). New York: Springer.

Wu, H., Lee, S., Chang, H., & Liang, J. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers & Education, 62,* 41- 49.

Yim, H. B., & Seong, P. H. (2010). Heuristic guidelines and experimental evaluation of effective augmented-reality based instructions for maintenance in nuclear power plants. *Nuclear Engineering and Design, 240*(12), 4096-4102.