

## Comparing Virtual and Location-based Augmented Reality Mobile Learning: Emotions and Learning Outcomes

**Jason M. Harley, Ph.D.** University of Montréal and McGill University

<sup>a</sup>University of Montréal, Computer Science and Operations, 2920 Chemin de la Tour, Pavillon André-Aisenstadt 2194, Montréal, QC, CAN, H3C 3J7

<sup>b</sup>McGill University, Department of Educational and Counselling Psychology, 3700 McTavish Street 614,<sup>[SEP]</sup>Montréal, QC, CAN, H3A 1Y2

[jason.harley@umontreal.ca](mailto:jason.harley@umontreal.ca)

Phone: (514) 561-3724

\*Corresponding author

*Bio:* Dr. Jason M. Harley is a postdoctoral fellow in the Computer Science and Operations Department at the University of Montréal and research associate at McGill University in the Department of Educational and Counselling Psychology. His research examines emotions, self and co-regulated learning, and advanced learning technologies, including intelligent tutoring systems, serious games, and mobile augmented reality applications.

**Eric G. Poitras, Ph.D.** University of Utah

Department of Educational Psychology, 1721 Campus Center Drive SAEC 3220 Salt Lake City, Utah 84112-8914

[eric.poitras@utah.edu](mailto:eric.poitras@utah.edu)

*Bio:* Dr. Eric G. Poitras is an Assistant Professor in the Instructional Design and Educational Technology program of the College of Education at the University of Utah. He is the director of the Advanced Instructional Systems and Technologies (ASSIST) laboratory, which aims to improve the adaptive capabilities of technology-rich learning environments across disciplines.

**Amanda Jarrell,** McGill University

McGill University, Department of Educational and Counselling Psychology, 3700 McTavish Street 614,<sup>[SEP]</sup>Montréal, QC, CAN, H3A 1Y2

[amanda.jarrell@mail.mcgill.ca](mailto:amanda.jarrell@mail.mcgill.ca)

*Bio:* Amanda Jarrell is a master's student in the Learning Sciences program at McGill University. She is a research assistant in the Advanced Technology for Learning in Authentic Settings (ATLAS) laboratory and a student of the Learning Environments Across Disciplines (LEADS) international research partnership.

**Melissa C. Duffy,** McGill University

McGill University, Department of Educational and Counselling Psychology, 3700 McTavish Street 614,<sup>[SEP]</sup>Montréal, QC, CAN, H3A 1Y2

[melissa.duffy@mail.mcgill.ca](mailto:melissa.duffy@mail.mcgill.ca)

*Bio:* Melissa C. Duffy is a PhD student in the Learning Sciences program at McGill University. Her research focuses on the role of emotion, motivation, and self-regulated learning processes within technology-rich learning environments.

**Susanne P. Lajoie, Ph.D.** McGill University

McGill University, Department of Educational and Counselling Psychology, 3700 McTavish Street 614,<sup>[SEP]</sup>Montréal, QC, CAN, H3A 1Y2

[susanne.lajoie@mcgill.ca](mailto:susanne.lajoie@mcgill.ca)

*Bio:* Dr. Susanne Lajoie is a Professor and Canadian Research Chair Tier 1 in Advanced Technologies for Learning in Authentic Settings in the Department of Educational and Counselling Psychology at McGill University and a member of the Centre for Medical Education. She is a Fellow of the American Psychological Association (APA), appointed for her outstanding contributions to the field of Psychology as well as a Fellow of the American Educational Research Association (AERA).

## **Comparing Virtual and Location-based Augmented Reality Mobile Learning: Emotions and Learning Outcomes**

*Abstract.* Research on the effectiveness of augmented reality (AR) on learning exists, but there is a paucity of empirical work that explores the role that positive emotions play in supporting learning in such settings. To address this gap, this study compared undergraduate students' emotions and

learning outcomes during a guided historical tour using mobile AR applications. Data was collected in a laboratory (Study 1;  $N = 13$ ) and outdoors (Study 2;  $N = 18$ ) from thirty-one undergraduate students at a large North American university. Our findings demonstrated that learners were able to effectively and enjoyably learn about historical differences between past and present historical locations by contextualizing their visual representations, and that the two mobile AR apps were effective both in and outside of the laboratory. Learners were virtually situated in the historical location in Study 1 and physically visited the location in Study 2. In comparing results between studies, findings revealed that learners were able to identify more differences outdoors and required less scaffolding to identify differences. Learners reported high levels of enjoyment throughout both studies, but more enjoyment and less boredom in the outdoor study. Eye tracking results from Study 1 indicated that learners frequently compared historical information by switching their gaze between mobile devices and a Smart Board, which virtually situated them at the historical location. Results enhance our understanding of AR applications' effectiveness in different contexts (virtual and location-based). Design recommendations for mobile AR apps are discussed.

*Mobile learning, augmented reality, virtual reality, emotions, history learning*

## **Introduction**

As technology becomes more ubiquitous and handheld devices more common, there is great potential for augmented reality (AR) approaches to foster learning *in-the-wild*. In other words, a mobile phone can be used to augment one's learning about the world around them, creating new and countless potential opportunities for informal learning as well as guided learning that takes place outside of the classroom.

Augmented Reality (AR; Azuma et al., 2001; Milgram, Takemura, Utsumi, and Kishino, 1994) promotes learning about the past through the delivery of digital media (i.e., sound, video, graphic, and text) that is sensitive to the specific exhibits and artifacts that are located in different historical and heritage sites. In other words, AR supplements reality (rather than replacing it like virtual reality; VR) with digital information designed to be relevant to the activity users are engaging in with an AR-supported device. In the case of history, this information could include texts about the history of a site, pictures of what it looked like in the past, an audio tour that describes a historical timeline, or a video

that provides a tour of historically-meaningful but difficult to access areas such as a bell tower or cellar.

In a recent review of AR in education Wu, Lee, Chang, and Liang (2013) identified a number of different definitions and taxonomies for understanding AR. According to the authors, our above description would be best categorized as a broad definition of AR rather than a restrictive one where real world context is dynamically overlaid with location or context-sensitive virtual information (Klopfer & Squire, 2008). Klopfer (2008) further characterizes AR in terms of the amount of digital media that is provided to the learner, ranging from lightly augmented reality where information is provided primarily through the real world setting, to heavily augmented reality where most of the information is provided virtually. AR technologies can be further differentiated in terms of sensors and devices. First, software applications utilize different types of sensors to capture either locations or objects in the real world. Location-aware or place-based AR utilizes global positioning system (GPS) or indoor position system (IPS) technology to track learners as they physically move throughout real world locations, in contrast to context-sensitive AR where 2D and 3D objects are scanned and recognized by an application. AR can also be implemented through several types of hardware, including desktop computers, handheld devices, and head-mounted displays that use markers and GPS amongst other technologies to dynamically augment user's information (i.e., to change text or image-based historical information based on a change in location; see Zhou, Duh, & Billinghamurst, 2008).

Research on the effectiveness of AR systems in fostering learning and positive emotions is in an early stage, and there is currently a paucity of empirical research with which to inform and guide the design of these systems (Wu et al., 2013). The studies of location-based AR in history education focus instead on the

development and usability of these systems (Ioannidis et al., 2011; Katifori et al., 2014; Keil et al., 2013; Rabina & Cocciolo, 2012; Vayanou et al., 2012). The research designs of such studies rely primarily on self-reports of usability and preference (Cocciolo & Rabina, 2013; Pujol et al., 2013), qualitative field observations (Pujol et al., 2013; Rennick-Egglestone et al., 2013), post-tour interviews (Cocciolo & Rabina, 2013; Pujol et al., 2013), and system generated log-file data (Vayanou et al., 2012), from small sample sizes, which are used to inform changes made to the design of the interface. As such, there is a lack of theory-driven empirical research that evaluates outcomes of AR applications with the purpose of iteratively improving their design guidelines to improve learning outcomes by fostering positive emotions, which are known to predict learning outcomes (Pekrun & Perry, 2014; see below for more on emotions and learning).

The implications of this line of research are therefore to establish evidence-based design guidelines for location-based learning scenarios that will support positive emotions and meaningful experiences in the context of learning about history. In the following sections of the introduction, we make initial efforts toward addressing this issue by: (1) reviewing prior AR research; (2) describing the location-based mobile AR applications and instructional scenario used in the studies this manuscript describes; (3) outlining and contextualizing our theoretical frameworks for history learning and emotions within the AR mobile environment used in the described studies, and (4) outlining the research questions and hypotheses that drove this study that evaluated the effectiveness of two applications described below (see *Augmented Reality Mobile Applications and Learning Environments*).

## **Related AR Learning Environments and Research**

AR has increased in popularity within the broader educational community, and is referred to as one of the key emerging forms of instructional technology (Martin et al., 2011). In a review of the educational literature, Wu et al., (2013) outlined several affordances of AR for educational purposes: (1) 3D virtual objects that learners can manipulate to observe and model physical phenomena; (2) explicit representations of otherwise invisible concepts and events; (3) collaborative instructional scenarios that are situated in real-life problems and applications; (4) a sense of community amongst learners that participate in comprehensive and realistic experiences; and (5) a direct connection between learning in formal and informal settings. The educational benefits associated with the multimedia content that is displayed in this medium includes but is not limited to collaborative problem-solving scenarios, narrative-driven inquiry-based simulations, and interactions with virtual agents. Location-based learning scenarios emphasize the affordances of mobile devices to deliver information about the physical environment learners are exploring, allowing them to collect evidence, or investigate issues (DeLucia, Francese, Passero, & Tortora, 2012; Klopfer, 2008; Rosenbaum et al., 2007). In doing so, AR can immerse learners into the past (as well as the present) by blending real-life settings, such as historical landmarks, with virtual information that can augment learning and engagement (Bronack, 2011; Broll et al., 2008; Klopfer & Squire, 2008).

Despite the potential of AR technologies for facilitating meaningful learning experiences, AR is seldom developed and utilized in the context of historical heritage sites and museum institutions (Bressler, 2013). Tallon (2013) surveyed the use of mobile devices by 551 museum professionals working in these institutions and related sectors across North American and 27 other countries. Although 43% of respondents reported that their cultural institution

offered mobile devices to visitors, the use of AR systems was less prevalent than audio guided tours or interactive experiences that incorporate social media.

There are a few noteworthy location-based AR systems that have been implemented in the context of historical tours to provide educational experiences. The Fog Altimeter application was released by the Exploratorium to allow users to view the Golden Gate Bridge superimposed with digital media, enabling users to measure the height of the fog near the bridge (Rothfarb, 2011). Another AR-based system is the House of Olbrich, which is designed to serve as an architecture guide for tourism on cultural heritage sites and provides interactive hotspots on buildings to provide detailed information about them (Keil et al., 2011). GeoStoryteller allows visitors to view archival photos and multimedia narratives during walking tours of historical sites in New York City. Visitors have the benefit of historic photographs that are layered against the real-life location as displayed through the mobile device camera. The multimedia narratives consist of instructional videos about the historic site and the lives of people that were involved in specific events. Trivia games are made available to visitors, involving short multiple-choice item assessments where the results can be shared through social media sites. The system relies on the Layar augmented-reality browser or a web-based delivery system to deliver topics in relation to the German immigration to New York City (Rabina & Cocciolo, 2013). A Creative Commons open-source application called GeoStories enables users to create their own narratives using location-sensitive Google maps to identify key GPS locations. The source materials are assembled from digitized resources from public institutions such as the New York Public Library and the Library of Congress.

Other AR systems implement interactive storytelling experiences to visitors inside the museum (Katifori et al., 2014; Vayanou et al., 2014). In this case, story authors, including curators, museum staff, and script writers, create

stories through an authoring tool called PAROS (Ioannidis et al., 2011), where story elements are pre-defined in accordance to conditional events and/or visitor characteristics of interest. Prototypical profiles of visitors are first created with the help of the tool, where at the beginning of the visit, a visitor is matched with the relevant profile, which is gradually refined through the collection of user interactions during the visit. In the Acropolis Museum for instance, visitors learn about the archaeological findings of the Acropolis of Athens, from the Greek Bronze Age to Roman times. Visitors walk amongst sculptural remains in the Archaic Gallery, located on the first floor of the museum. The archaeological objects are organized in accordance with themes and chronology, which is indicated to the visitor through labels and panels. One of the stories depicted by the augmented reality application refers to a horse as the narrator of the guided walking tour. The topic of the story consists of the significance of horses to ancient Greeks, spanning from daily life activities to mythology. The plot develops over time, relying on reference points within the exhibit to determine its progress. These examples reflect the diversity and opportunities for creativity that AR applications can afford history learning in both formal and informal education.

### **Augmented Reality Mobile Applications and Learning Environments**

In the two studies this paper examines, we used two mobile applications to augment learners reality: one was managed by the learner (MTL Urban Museum) and the second by the (human) tour guide (MetaGuide). The human tour guide was either a postdoctoral fellow or undergraduate research assistant (varied by study). A single guide interacted with a single learner per experimental session. The MTL Urban Museum app is an app developed by the McCord Museum that allows users to locate historical landmarks in the city of Montreal using GPS



smart phone technology and to view multimedia content (text and images) that illustrate how the city used to look in the 1800s in comparison to the present day. Figure 1 shows the app's interface after a location (The Roddick Memorial Gates, used in the discussed studies) has been selected from the map of the city (populated by pins corresponding to locations the app has multimedia content for). A second interface (not displayed) includes a textual description of the Roddick Memorial Gates (referred to henceforth as the Roddick Gates) that appears when the historical image is clicked on. Figure 2 shows a learner making historical comparisons between the picture of the Roddick Gates displayed on the MTL Urban Museum app and a virtual representation of the present day version of the Roddick Gates displayed on a SMART Board using Google Earth (Study 1; Study 2 took place on location of the Roddick Gates and therefore did not necessitate virtually situating learners at the historical site).

--Insert Figure 1 about here--

--Insert Figure 2 about here--

The MetaGuide app was developed by the second and fifth authors of this manuscript to be used in conjunction with the MTL Urban Museum to support historical inquiry (described below in Theoretical Framework for Fostering Historical Reasoning section). Unlike the MTL Urban Museum app, which was handled by the learner, MetaGuide was primarily used by the guide. Figure 3 illustrates a screenshot of the primary MetaGuide interface where three main (superordinate) historical differences are outlined in red to remind the guide of the learning objectives for the learning session (for the learner to point out differences between them in the 1800s vs. the present day). Figure 4 displays a visualization

of specific historical differences for each of the three superordinate ones that learners could point out: (1) differences about Buildings (i.e., contemporary absence of the Porter’s House), (2) Transportation (i.e., contemporary decline in popularity of horse drawn carriages as a mode of Transportation), and (3) the Gates themselves (i.e., material: picket fence vs. stone pillars). Figure 5 displays an example of a prompt that MetaGuide would provide the guide with to use in order to direct a learners’ attention to a specific historical difference or evaluate whether the difference a learner pointed out is correct. Guides could access prompts for each of the superordinate historical differences by touching the red outline on the main interface (Figure 3). Prompts vary depending on the historical difference in question, and include (for example) an illustrative timeline of changes to the Roddick Gates (if the Gates are selected) and an illustration of the porter’s house (if the left-most red area is selected), which was mentioned in the text. Guides also possessed a hard copy of the prompts (including the historical illustrations) available in MetaGuide as part of the detailed experimental protocol they kept on hand and consulted during the experiment.

--Insert Figure 3 about here--

--Insert Figure 4 about here—

--Insert Figure 5 about here--

Table 1 provides a summary of Studies 1 and 2, including similarities and differences; chief amongst them, that Study 1 was conducted in a laboratory setting that virtually simulated the historical location and Study 2 took place *at* the historical location (in person). The two apps were used in both studies.

-- Insert Table 1 about here --

The virtual guided tour (Study 1) was supported with the Google Earth Street View program displayed on a SMART Board. Learners were virtually positioned in front of the historic location (the Roddick Gates) at McGill university and were invited to navigate from one place to another by changing the orientation of their view through 360-degree panoramic imagery of supported locations (Google Earth) displayed on the large, touch sensitive SMART Board. Learners met the experimenter at the historical location in-person in Study 2. Learners used the MTL Urban Museum mobile app to compare historic images of the location with the present in both Studies 1 and 2. Learners navigated through the MTL Urban Museum's map of [X] in order to locate the Roddick Gates in relation to their present (or virtual) location using GPS technology in both studies.

Compared with previous AR environments and research, Study 1 represented a mixed reality environment that contained more virtual than real world features and objects, and would accordingly be classified on Klopfer's (2008) continuum as a more heavily augmented experience than Study 2 where learners were physically rather than virtually situated at the historical location. Learners tours were lightly augmented in both studies with historical multimedia information provided in the apps, including the location-aware MTL Urban Museum app (i.e., GPS).

In the following section we describe specific theoretical frameworks used to inform the design of this environment (differences and similarities outlined in Table 1) and hypotheses for the current studies which help to make them unique amongst research done with AR. In other words, using theory to frame the design of mobile AR apps is novel. In the absence of an integrated, domain-specific

theoretical framework we drew upon several different theoretical frameworks that provided recommendations for designing learning environments that would support historical reasoning, positive emotions, adaptive cognitive processing and adaptive learning behaviors.

### **Theoretical Framework for Fostering Historical Reasoning**

We see the potential of AR as an agent for helping learners construct their own meaning of history as they explore historical artifacts. Our theoretical framework is guided by van Drie and van Boxtel's historical reasoning framework (2008). This framework as well as the theoretical frameworks presented in the proceeding section on emotions and cognitive processing informed the design of the study by providing insight on how to support different adaptive processing during participants' interactions with the apps (learning about history, experiencing positive emotions, avoiding cognitive overload). Historical reasoning is operationalized as an activity during which learners acquire knowledge and interpret facts about the past. One way learners engage in historical reasoning is by comparing or explaining historical phenomena. In this regard, learners adopt an active role in constructing knowledge of the past. The framework of historical reasoning proposed by van Drie and van Boxtel (2008) contains six components: (1) posing historical questions, (2) using sources of information, (3) contextualization, (4) argumentation, (5) using substantive concepts and (6) using meta-concepts (van Drie & van Boxtel, 2008). Our work incorporates all six of the core components of this historical reasoning framework. A description of these components and their implementation into MetaGuide and the MTL Urban Museum is described below.

### *Posing Historical Questions.*

The ability to recognize and understand historical questions is an essential component of historical reasoning (Schreiber et al., 2006). One type of question that can prompt historical lines of reasoning is comparison questions (van Drie & van Boxtel, 2008). MetaGuide prompts the guide to ask learners to compare past and present locations in order to identify differences. These open-ended comparative questions foster historical reasoning by encouraging learners to explore how locations have changed from the past to the present.

### *Using Sources of Information*

Learners can acquire information about the past from written documents and images. Reasoning with historical documents involves using information presented in documents or images (Rouet et al. 1996). MetaGuide and the MTL Urban Museum provide learners with both written text and images to give information about the historical location they are visiting (in person or virtually). Learners must reason with this multimedia information in order to answer the comparative questions posed by the guide.

### *Contextualization*

To interpret and understand the past learners must apply general knowledge about characteristics of the time period and place (van Drie & van Boxtel, 2008). Learners must situate historical phenomena or objects in a spatial and social context in order to compare events from the past to the present. MetaGuide and the MTL Urban Museum facilitate contextualization by providing information about the historical date, location and situating learners in the present-day location (i.e., horse drawn carriages no longer being used a mode of Transportation).

### *Argumentation*

Historical information is often partial and contradictory because historical representations are not perfect or absolute. Making claims about the past must be supported by rational arguments and evidence (van Drie & van Boxtel, 2008). MetaGuide and the MTL Urban Museum implicitly foster argumentation by providing partial information about the site location. When participants were making claims about the past using the sources of information provided, they must implicitly form an argument to justify their response. For example, a participant who notes that there were more horse drawn carriages in the 1800s might justify their claim using evidence from the image (i.e. the image depicted a horse drawn carriage and no cars) or using knowledge about that time period to form a rational argument (e.g. there were more horse drawn carriages because cars were too expensive for the average citizen to purchase).

### *Using Substantive Concepts*

Concepts such as historical phenomena, structures, persons and periods enable learners to thematically organize historical knowledge (van Drie & van Boxtel, 2008). The meaning of concepts, however, differs according to time and place. Individuals must learn how to differentiate between the present meaning of a concept and the historical meaning. MetaGuide and the MTL Urban Museum application encourage the use of substantive concepts by asking learners to identify differences between the past and present. For example, present day Transportation is different from Transportation in the 1800s. To make comparisons between Transportation today and Transportation in the 1800s, participants must calibrate the meaning of the concept to be appropriate for the time period.

### *Using Meta-concepts*

Meta-concepts include evidence, time, place and change (van Drie & van Boxtel, 2008). Meta-concepts are important for historical reasoning because they guide questions about historical comparisons. MetaGuide and the MTL Urban Museum applications allow learners to make comparisons and assess change over time. In order to foster historical reasoning, another key consideration is the affective states that learners experience while interacting with the learning environment. In the following section we discuss how learners' emotions relate to features of the learning environment, learning processes, and outcomes.

### **Theoretical Framework for Fostering Positive Affect and Adaptive Cognitive Processing**

According to control-value theory of achievement emotions (Pekrun, 2006; Pekrun, Frenzel, Goetz, & Perry, 2007; Pekrun & Perry, 2014) learners' emotional states (both positive and negative) are predicted by subjective appraisals of control and value (i.e., proximal antecedents of emotion). For instance, if a learner feels highly in control of the task and also highly values a task, they are expected to experience enjoyment. On the other hand, if they do not value the task, they will likely experience boredom irrespective of their appraisals of control. These emotional states are critical because according to this theory and a host of empirical work (for review see Pekrun & Perry, 2014), emotions are expected to predict learning outcomes through their influence on motivational, cognitive, and metacognitive processes. For instance, high degrees of enjoyment may free up cognitive resources given that attention is directed toward the task at hand, which allows learners to engage in complex learning processes (i.e., historical reasoning). On the other hand, high degrees of boredom may direct

learners' attention away from the task at hand, which hinders learning (Pekrun, 2006; Pekrun et al., 2007; Pekrun & Perry; 2014).

Previous research has indicated that positive activating emotions (i.e., high physiological arousal and positive-valence state), such as enjoyment, are typically related to adaptive learning outcomes, whereas negative deactivating emotions (i.e., low physiological arousal and negative-valence state), such as boredom, are typically related to maladaptive learning outcomes (Pekrun & Perry, 2014). For example, Pekrun, Daniels, Stupnisky and Perry (2010) found that boredom was positively related to attention issues and negatively related to intrinsic motivation, self-regulation, effort and performance. Other studies have found similar maladaptive trends (e.g., Mann & Robinson, 2009; Pekrun, Hall, Goetz, & Perry, 2014; Watt & Vodanovich, 1999). As such, it is critical that research examines whether learning environments promote more positive activating emotions. In this study we focused on activity emotions assessed before or after their learning session (emotions directed toward the task or activity at hand [e.g., enjoyment and boredom]), rather than outcome emotions (emotions directed toward success or failure outcomes [e.g., pride, shame]) given that learners were asked to participate in a learning activity without any strong performance or achievement indicators.

An important implication of the control-value theory is that distal antecedents of emotions, including features of the learning environment (e.g., content, tasks, medium), can be designed to enhance learners' control and value appraisals, which in turn will lead to more positive emotions and facilitate adaptive learning processes and learning behaviors (Pekrun, 2006; Pekrun, et al., 2007; Pekrun & Perry; 2014). For instance, the MTL Urban Museum app affords user-directed navigation, which provides learners with more choice in what they attend to, for how long, and in what order. These features can enhance learners' perceptions of control by supporting autonomy and self-directed inquiry. In terms



of enhancing value, the MTL Urban Museum App is designed to provide interactive and dynamic learning through multiple media representations (e.g., text and images) and heightened realism, which can enhance interest. It is also designed to increase value by providing historical information about a real-world setting that is (in the studies described in this manuscript) personally relevant to learners (a structural and symbolically important part of the university they attend), which can enhance utility. Collectively, these features are expected to foster enjoyment by increasing perceptions of control and value.

This is also consistent with Mayer's multimedia learning theory (2005; 2009; 2014a; 2014b), which posits that incorporating emotionally and motivationally appealing design features (e.g., attractive content, graphics) can help to increase cognitive engagement and retain learners' attention. At the same time, Mayer (2009) argues that it is important to incorporate design features in the learning environment that reduce extraneous processing (e.g., eliminate distracting content, unnecessary text) as this can lead to task-irrelevant thinking by taxing working memory and directing attention away from the essential processing needed to achieve learning goals. Accordingly, the multimedia MTL Urban Museum App was designed to reduce extraneous processing for learners to make historical comparisons by using a single historical picture of the Roddick Gates with a small amount of text that can be viewed and made invisible, in turn, by tapping on the screen. In the following section we elaborate on the goals of the current study and outline our hypotheses.

## **The Current Studies**

The aim of these studies was to evaluate the effectiveness of two mobile-based applications, used together, to foster positive emotions and meaningful learning in a laboratory (Study 1; lab-based, virtually-situated location) as well as

an outdoor, location-based study (Study 2; outdoor, location-based). This study advances the state-of-the-art of AR by focusing on the learning and affective outcomes of student interactions with AR in both virtually-situated and on-site historical locations. Moreover, behavioral process data (i.e., eye-tracking) was used to examine learners use of multiple sources of historical information in the virtual study (Study 1). The theoretical framing of the mobile AR apps and this study are also novel to the field.

The following research questions were examined in Studies 1 and 2 in order to evaluate the AR mobile apps in terms of their collective ability to facilitate historical contextualization and positive emotions: (1) Were learners able to identify historical differences between the past and present? (2) Did learners report enjoying their learning session? To address these questions, this study relied on a novel methodological approach that combines affordances of a laboratory and field setting. Study 1 was conducted in a laboratory setting to virtually re-create the historical setting, whereas Study 2 took place in the outdoor location. We explored a (3) third research question that was unique to the laboratory-based study where a mobile eye-tracker (described below) was used to answer the following question: Did learners use the different AR and mobile features of the virtual tour (i.e., mobile devices displaying apps vs. SMART board displaying Google Earth) to help them make comparisons and provide additional historical context? This last question allowed us to assess learners *use* of the apps provided to them to help engage in historical reasoning (Rouet, et al., 1996; van Drie & van Boxtel, 2008). Accordingly, we were particularly interested in how often learners made comparisons between information presented on the different devices (i.e., apps). Given that the information presented on the different devices was historical in nature we interpreted gaze transitions between devices as an indicator of historical reasoning, specifically, making historical comparisons.

Previous research has demonstrated a relationship between where the eye is looking and cognitive processing (e.g., eye-mind assumption; Just & Carpenter, 1980; Mason, Tornatora, Pluchino, 2013; van Gog, 2014).

We hypothesized that the apps would effectively promote history learning since the apps were designed to promote inquiry using the six components of the historical reasoning framework (van Drie & van Boxtel, 2008). These components were: posing historical questions, using sources of information, contextualization, argumentation, using substantive concepts and using meta-concepts. We also hypothesized that the apps would promote positive emotion and learning behaviors since the features of the AR learning environment are designed to enhance learners' perceptions of control and value (e.g., user-directed, multi-media, high realism, personally relevant). According to the control-value theory of achievement emotions, autonomy-supportive and meaningful learning environments are expected to predict positive emotions (e.g., enjoyment), and in turn, adaptive learning behaviors. Consistent with Mayer's multimedia learning theory, we predicted that enjoyment would be higher in the outdoor setting (Study 2) compared to the lab setting (Study 1) given that the real-world environment may reduce cognitive load by reducing discrepancies between the multiple representations of the setting. In addition, we would also expect that a real-world setting would heighten relevance and realism, which may increase value and lead to greater enjoyment.

## **Methods**

### **Participants**

In Study 1, thirteen undergraduate students (6 female; 7 male) from a large North American university volunteered to participate. Participants were between 18 and 35 years old ( $M = 24.38$ ;  $SD = 4.72$ ) and enrolled in various programs

(three from science or engineering; two from business-related fields; and eight from bachelor of arts programs, including one history major). Student GPAs ranged from 2.4 to 3.92 ( $M = 3.33$ ;  $SD = 0.47$ ) out of four. There was no preference for participants' academic background other than enrollment as a full time undergraduate student at X university, and therefore no rationale for controlling demographics. Participants were compensated with \$5 at the end of the session and recruited from the university's online classified ad in both studies. Eligible participants were scheduled on a first contact-first schedule basis (we had more interest than available timeslots due to limited human and financial resources for these studies). All scheduled participants were run if they were (a) available to participate during one of the scheduled experimental sessions (all sessions scheduled between the hours of 9AM-6PM on weekdays) and (b) showed up for their session.

In Study 2, eighteen undergraduate students (15 female; 3 male; 1 unreported) from the same large North American university volunteered to participate. Participants were between 18 and 26 years old ( $M = 20.72$ ;  $SD = 2.29$ ) and enrolled in various programs (11 from science or engineering; 3 from business-related fields; and 4 from bachelor of arts programs). Student GPAs ranged from 2.22 to 4.00 ( $M = 3.37$ ;  $SD = 0.48$ ) out of four.

## **Measures and Materials**

The augmented reality (AR) environment consisted of two mobile applications described above: one for the learner (MTL Urban Museum) and another for the tour guide (MetaGuide). The laboratory (Study 1) provided opportunities for collecting data on behaviors, learning, and emotions before, during, and after the tour, including: portable eye-tracking glasses by SensoMotoric Instruments (SMI: ETG 2w; 30Hz) to monitor learners attention to:

a (1) smart phone displaying the MetaGuide app, (2) a smart phone displaying the McCord MTL Urban Museum app, as well as (3) a SMART Board displaying the Google Earth visual of the Roddick Gates (SMART Board); audio-recordings of verbal interactions between the learner and the guide; and self-report measures of enjoyment and boredom.

In order to ensure the accuracy of the portable eye-tracking results we conducted a 1-point calibration procedure for each participant prior to commencing eye-tracking recording. We verified the accuracy of the calibration by instructing the participant to focus their gaze on a specific word presented on the smart phone for approximately 10 seconds while the eye-tracker was calibrated. During this process, the eye-tracker captured the unique characteristics of the users' eyes to track gaze data. Calibration was therefore established using an area of interest at a finer level of granularity than necessary for the analyses conducted (e.g., SMART Board, smart phones), ensuring a high degree of accuracy for the current analyses and the opportunity to conduct more granular analyses (e.g., specific words, portions of images) in the future. Moreover, eye-tracking recordings were monitored in real time by one of the experimenters (not the tour guide) throughout each participant session to verify accuracy. If inaccuracies were detected, a recalibration could be conducted during the session. Calibration corrections could also be conducted post-hoc to address any inaccuracies or change in calibration during recording, though none were needed post-hoc. Audio data was recorded using the portable eye-tracking device in Study 1.

Two *self-report measures* assessed learners' level of enjoyment. The first was a three-item questionnaire administered before the learning session and asked learners how much they anticipated enjoying the virtual tour, learning about historical differences, and interacting with the guide. After the learning session

participants answered a six-item version of the measure, which asked them about their level of enjoyment again as well as how bored they felt for each component (Guide, Tour, Learning; see Table 5 in the Results section). The questionnaire was an adapted version of the academic achievement emotion questionnaire (Pekrun, Goetz, Titz, & Perry, 2002; Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011) and used a five-point Likert scale.

Students also answered two open-ended questions at the end of the experiment: (1) *What did you like about the McCord Urban Museum iPhone application?*; and (2) *What did you dislike about the McCord Urban Museum iPhone application?* Students' answers were analyzed for spontaneous and voluntary emotional statements to provide an additional measure of post-tour emotions.

In the second study, the measures and materials used were identical to the first study except that eye-tracking data was not recorded, audio data was recorded using the Voice Memos program (rather than integrated eye-tracker audio recorder) on a lab iPhone, and a SMART Board was not used. Instead, learners met the experimenter in front of the historic location itself (the Roddick Gates) and made comparisons between them and the digital image of the location on the MTL Urban Museum App (standing in the same location as they were virtually positioned in Google street view). The questionnaires were filled out on the iPhone rather than on a computer.

## **Experimental Procedure**

The laboratory study (Study 1) took approximately 25 minutes. Participants completed an on-line demographics questionnaire prior to participating to ensure they met study criteria (undergraduate student). On the day of the study, participants first completed a consent form and the self-report

pretest. Participants were then calibrated to a portable eye tracker by an experimenter and shown how to use the interactive SMART Board and Google Earth application (using the street view) to navigate. Participants were instructed on how to locate the Roddick Gates using the mobile app and once they located the gate they read a short text about the Porter house. Participants were then invited to identify differences between the 1800s version of the Roddick Gates displayed on their smart phone and the contemporary Gates viewable on the interactive SMART Board.

In contrast, the outdoor, location-based study (Study 2) had duration of approximately 10 minutes. Participants completed an on-line demographics questionnaire prior to participating to ensure they met study criteria (i.e., were undergraduate students). On the day of the study, they first completed a consent form and the self-report pretest. Participants were then invited to identify differences between the 1800s version of the Roddick Gates displayed on the smart phone they were using the MTL Urban Museum on and the contemporary Gates in front of them. Learners were positioned in the same position relative to the Roddick Gates as they were virtually (a few meters in front of them; see Figure 2). While learners were invited to change their location relative to the Roddick Gates, participants tended to remain stationary and modify their orientation (e.g., physical or virtually-simulated head position), rather than walk a physical or virtual path. Learners were invited to orient their attention and physical or virtual orientation/movement in a sequence that they chose.

## **Data Analyses**

To answer the *first research question*, learners' dialogues with the guide were coded to determine if they could identify historical differences between past and present (see Table 2). Historical differences stemmed from three broad

superordinate categories outlined in Figure 4 (Gates, Buildings, and Transportation). The coding scheme and exemplar codes presented in Table 2 differentiate between historical differences that undergraduate student learners were able to identify as well whether learners had help identifying them. The coding scheme differentiates superordinate historical differences (e.g., between 1800 versus present day Transportation; see Figure 4) learners were able to identify independent of any prompts or information provided by the guide versus differences the guide pointed out to (i.e., told) the learner. At the subordinate level (e.g., carriage for Transportation; see Figure 4) the coding scheme differentiated situations where learners identified historical differences with help (scaffolding) from the human tour guide in addition to differences that were identified independently by the learner or pointed out by the guide. This coding scheme yielded the data we used to examine learning outcomes discussed in the results section (frequencies of historical differences identified independently, with scaffolding, or by the guide: (1) over-all, (2) by superordinate difference, and (3) by subordinate difference). Two experimenters coded 73 segments (from 12 of the 13 participants) of dialogue using this coding scheme with an agreement rate of 90.4% ( $\kappa = .85$ ) for mutually coded segments in Study 1. For Study 2, two experimenters coded 125 segments (from 16 of the 18 participants; one participant was used for training coders and another participant's audio recording data was not captured and therefore unavailable) of dialogue using the coding scheme presented in Table 2 with an agreement rate of 96.8% ( $\kappa = .94$ ).

The guide's prompts and feedback to learners' comments about historical differences were coded as well (see Table 3). For the purposes of this study, this coding scheme was used to help classify the guide's discourse with the learner to ensure understanding between coders regarding which guide prompts were considered (1) providing scaffolding, (2) answers to students, or (3) other types of



discourse (e.g., clarifications and agreement) not immediately relevant to the examination of learning through the identification of historical differences. Two experimenters coded 131 segments (from 12 of the 13 participants) of dialogue using this coding scheme with an agreement rate of 85.5% ( $\kappa = .83$ ), and a post-discussion agreement of 100%. In Study 2 two experimenters coded 246 segments (from 16 of the 18 participants) of dialogue using this coding scheme with an agreement rate of 91.9% ( $\kappa = .87$ ) for mutually coded segments, and a post-discussion agreement of 99.2%.

---Table 2 about here ---

---Table 3 about here ---

We examined learners' scores of self-reported enjoyment toward different object foci (guide, learning, tour) to answer the second question (did learners report enjoying their learning session?). Eight statistical outliers (out of 261 data points) were identified using SPSS's boxplot analyses. These outliers were identified by examining the distribution of responses to each emotion question (nine) on the PreTest and PostTest separately for each study, given that between group comparisons were made in the analyses (Tabachnick & Fidell, 2007). Outliers included: one value (i.e., data point) in the PostTest questionnaire on learning-related enjoyment in Study 1; two in the PostTest questionnaire on tour-related boredom in Study 2; three in the PostTest questionnaire on learning-related boredom in Study 1; and two for guide-related boredom in Study 1. No outliers were identified on PreTest emotion items. Only two of the eight outlying values (tour and guide-related boredom) came from the same participant, and given the number of remaining, non-outlying emotion values (seven of nine) for

this participant, it was decided to retain them in the sample. Outliers were replaced with the next most extreme score (Osborne & Overbay, 2012; Tabachnick & Fidell, 2007). Data from two of the 18 participants were lost in Study 2 due to a technical error submitting the self-reports online (Wi-Fi was unreliable outside).

Students' responses to two open-ended questions (asked at the end of the emotion questionnaire: (1) *What did you like about the McCord Urban Museum iPhone application?* and *What did you dislike about the McCord Urban Museum iPhone application?*) were analyzed for spontaneous emotional statements to provide an additional measure of post-tour emotions. These statements are defined as spontaneous because participants were not explicitly asked to report their emotional experience in response to the guide, learning or tour. Two human coders (also authors of this paper) reviewed all twenty-nine (two participants' questionnaire data was lost) volunteered statements and identified statements that related to the experience of a positive or negative emotion pertaining to the guide, learning or tour. Both coders drew upon their expertise on emotions, including the wording of questionnaires (e.g., AEQ) and definitions of emotions (Harley, Bouchet, Hussain, Azevedo, & Calvo, 2015; D'Mello, Lehman, & Person, 2010; Pekrun & Perry, 2014). Only statements that described an emotional experience were identified for coding. All other sections that did not describe an emotional experience were excluded from the analysis.

Coders classified participants' spontaneous emotional statements into one of 4 emotion categories: positive (e.g., enjoyment, happiness, curiosity, hope; PNO18: "love history and I love McGill so learning more about both is exciting to me"), negative (e.g., anxiety, boredom, frustration, hopelessness; PNL5 "...it definitely made me feel nervous"), absence of negatively-valenced emotion (e.g., lack of anxiety, boredom; PNO1: "...not to bore people with long text") or

absence of positively-valenced emotion statements (e.g., lack of enjoyment, curiosity; there were no instances of this type of emotion statement). .

Of all participant responses, 52% described an emotion experienced in response to the first question (i.e. 15 statements) and 24% described an emotion experienced in response to the second question (i.e. 7 statements). The two experimenters independently coded a total of 21 segments of participant responses. There was an agreement rating of 93% ( $\kappa = .72$ ) for the first question and 71% ( $\kappa = .52$ ) for the second. A post-discussion agreement of 100% was reached for both questions. This discussion resulted in one statement from the second question being identified as non-emotional. Thus, the final agreement for the second question was based on statements from 21% of participants (i.e. 6 statements).

We examined eye-tracking behavior to answer the *third question* (did learners use the app-based features designed to help them make historical comparisons and provide additional context?) to determine how often they looked between the interactive whiteboard (SMART board) Google Earth Display and their mobile device, which displayed the MTL Urban Museum app to identify historical differences. We also coded transitions between either of these objects (i.e., areas) of interest and the guide's mobile app, which displayed the MetaGuide app. Given our research question, which focused on examining the usage of different devices' applications (as opposed to what regions of these apps or sections of images therein participants were attending to) it was not necessary to examine more fine-grained areas of interest (AOIs). AOIs (mobile devices, SMART board) were manually coded by a research assistant after the tour and experimental session was complete using the exported video replay with the gaze behavior overlay as seen in the screenshot presented in Figure 2). In this example, the participant was coded as looking at the smart phone where the MTL Urban

Museum app was displayed. Gaze transitions between the aforementioned objects of interest were the only eye-tracking data coded and analyzed in this study although different metrics (i.e., gaze fixation duration) and finer-grained analyses exist (i.e., saccades) (Bondareva et al., 2013; Jaques, Conati, Harley, & Azevedo, 2014). Given the dynamic nature of the environment and the small number of areas of interest for this study, we opted to manually code gaze transitions rather than use gaze-mapping software. We found this approach to be feasible and efficient.

## Results

### Study 1 Findings: Laboratory Setting

#### *Question 1: Were learners able to identify historical differences?*

About 85% of learners independently identified at least one general (superordinate; Gates, Buildings, Transportation) historical difference. Historical differences between the Roddick Gate itself were identified by 69% of learners, while differences in Transportation and surrounding Buildings were identified by 23% and 15% of learners, respectively.

With regard to specific (subordinate; see Figure 4) historical differences (Table 4): 69% of learners independently identified at least one specific historical difference (e.g., for Gates: change from picket fence to pillared gate). The guide supported learners identification of at least one specific historical difference through scaffolding for all learners and provided an answer for 92% of learners. There was no set number or limit on the amount of differences that could be detected. Table 4 reveals that learners identified the greatest number of differences (38) for the Gate, both overall and independent of support from the guide (16; 54%). Most learners (77%) received scaffolding to identify historical differences between the past and present versions of the Gates and all received

scaffolding for identifying historical differences between modes of Transportation (16 instances across 13 participants) in contrast to only about half of the participants receiving scaffolding to identify differences relating to the Buildings (54%). The greatest number of differences noted between past and present was of the Gates themselves (38) in comparison to those identified for Buildings (23) and Transportation (24).

Participants spent most of their time ( $M = 48.65\%$ ;  $SD = 8.04$ ) discussing differences between the two versions of the Gate in comparison to about 32% ( $SD = 11.15$ ) and 20% ( $SD = 9.53$ ) on Building and Transportation differences, respectively. Common historical features that participants identified included the changes in the shape and material of the Gates, the absence of horse drawn carriages as a means of Transportation, and the absence of extensions to the arts Building. However, only 15% of participants noted the absence of the clock tower in the 1800s and identified the absence of the porter's house in the contemporary version of the Gates.

---Table 4 about here ---

### *Question 2: Did learners report enjoying their learning session?*

The results of the pre and post-test emotion self-report measure (related to question 3) are reported in Table 5. Learners reported relatively high mean anticipatory and retrospective levels of enjoyment as well as low retrospective levels of boredom toward each of the object foci: the tour, learning, and the guide.

---Table 5 about here ---

The analyses of students' open-ended responses revealed that in response to the first question "*What did you like about the McCord Urban Museum iPhone application?*" 6 participants (i.e. 46% of all participants) made relevant statements. Examples of positive emotional states include: (PNL09) "*It's interactive, fun and easy to use. Contains interesting information.*"; (PNL13) "*I enjoyed the tour guide input of the application. As a McGill student, who walks past the Gates on a daily basis, the additional information makes me appreciate my surroundings far more than just it being the entrance to McGill.*" In addition, one participant (i.e. 8%) made a statement associated with experiencing the absence of a negative emotion: (PNL6) "...not to bore people with long text."

In response to the second question "*What did you dislike about the McCord Urban Museum iPhone application?*" two participants (i.e. 15%) reported experiencing a negative emotion. Examples include: confusion (PNL10) "*The tour guide showed me other pictures but it was unclear as to whether I had additional pictures within my application*"; anxiety (PNL5) "*I felt like maybe I was giving the wrong answers...definitely made me feel nervous.*" In addition, one participant (i.e. 8%) made statements that expressed positive emotions: (PNL1) "*It was pretty interesting and fun.*"

All remaining participants did not volunteer emotion-related statements, even if they were favorable toward the study (e.g., (PNL7) "*comparison with modern view of the Roddick gates allow[ed] me to reflect control over experience*"; (PNL4) "*I liked that you can see the pictures at different years*"). In addition 11 of the 13 participants provided constructive suggestions for improving the app in their responses to the second question, including increased automaticity and interactivity, and a desire for more content (discussed in the Future Directions).

*Question 3: Did learners use the different features of the TRE to help them make comparisons and provide additional historical context?*

Table 6 provides information about learners gaze transition between Google Earth and the mobile applications (learners and the guide's). On average, learners spent four-minutes and thirty-four seconds making historical comparisons by switching their gaze between the interactive whiteboard and their own or the guide's mobile devices about 48 times (20 times per minute).

---Table 6 about here ---

## **Study 2 Findings: Field Setting**

*Question 1: Were learners able to identify historical differences?*

About 94% of learners independently identified at least one general (superordinate) historical difference. Historical differences between the Gates themselves were identified by 88% of learners, while differences in Transportation and surrounding Buildings were identified by 47% and 35% of learners, respectively.

With regard to specific (i.e., subordinate; see Figure 4) historical differences: 88% of learners independently identified at least one specific (subordinate) historical difference (Table 4). The guide supported learners identification of at least one specific historical difference through scaffolding for 94% of learners and provided an answer for 77% of learners. There was no set number or limit on the amount of differences that could be detected. Table 3 reveals that learners identified the greatest number of differences (37) for the Gate both over all and independent of support from the guide (82%). Consequently, participants received the least amount of scaffolding for identifying historical differences between the past and present versions of the Gates. Learners received

the most scaffolding for identifying historical differences between modes of Transportation (18 instances across 13 participants) and the most answers from the guide for differences regarding Buildings (13 instances across 13 participants). The greatest number of differences noted between past and present was of the Gates themselves (37) in comparison to those identified for Buildings (32) and Transportation (31).

Participants spent most of their time ( $M = 44.10$ ;  $SD = 8.04$ ) discussing differences between the two versions of the Gate in comparison to about 37% ( $SD = 9.25$ ) and 27% ( $M = 19.38$ ;  $SD = 8.28$ ) on Building and Transportation differences, respectively. Common historical features that participants identified included the changes in the shape and material of the Gates, the absence of horse drawn carriages as a means of Transportation, and the absence of extensions to the arts Building. However, only 12% of participants noted the absence of the clock tower in the 1800s and 29% identified the absence of the porter's house in the contemporary version of the Gates.

### *Question 2: Did learners report enjoying their learning session?*

The results of the pre and post-test emotion self-report measure (related to question 3) are reported in Table 5. Learners reported relatively high mean anticipatory and retrospective levels of enjoyment as well as low retrospective levels of boredom toward each of the object foci: the tour, learning, and the guide. In response to the first question “*What did you like about the McCord Urban Museum iPhone application?*” the analyses of students’ open-ended responses revealed that seven participants (i.e. 44%) used words associated with the experience of a positive emotion. Examples include: (PNO2) “*It's neat that it can determine your location.*”; (PNO9) “*It was really cool to see all of the historical photos of a modern landmark I am quite familiar with*”; (PNO18) “*It was very*



insightful and interesting. I love history and I love McGill so learning more about both is exciting to me.” There was 1 participants that reported a lack of a negative emotion (i.e. 6%): (PNO01) “...without feeling rushed”).

In response to the second question “*What did you dislike about the McCord Urban Museum iPhone application?*” two participants (i.e. 13%) reported experiencing a positive emotion: (PNO2) “...interesting if you feel like learning something new.”; (PNO4) “...would like to know more about other buildings surrounding.” Only one participant reported experiencing a negative emotion (PNO13): “I was not interested in learning about the Roddick gates so I didn't care.”).

All remaining participants did not volunteer emotion-related statements, even if they were favorable toward the study (e.g., (PNO04) “It's pretty! Able to zoom and see closer is nice.”; (PNO10) “...simple to understand, good for a short overview. I liked the different pictures...”). In addition, seven of the 16 participants provided constructive suggestions for improving the app in their responses to the second question, most centering on a desire for more content (discussed in the Future Directions).

## **Comparative Findings: Study 1 and 2**

### *Question 1: Were learners able to identify historical differences?*

Differences in learners ability to identify historical differences between Studies 1 and 2 are observable in Table 4. Specifically, the results reveal that participants were able to identify more superordinate and subordinate (i.e., specific) differences in the outdoor, location-based study than the lab-based one, and that they also required less scaffolding and fewer answers being provided by the guide in order to identify key super and subordinate differences between the past and present versions of the Roddick Gates. A greater percentage of

participants identified the porter's house as absent from the current Roddick Gates in the outdoor study as well, despite participants from both groups reading a short text about it before beginning to identify historical differences.

#### Question 2: Did learners report enjoying their learning session?

Repeated measure ANOVAs were run to examine whether any significant differences existed between pre and post test scores of enjoyment as well as whether interaction effects existed for study and time. There was a significant effect of administration of the emotion questionnaire for tour-related enjoyment ( $F(1,27) = 11.73, p < 0.05, \eta^2_p = .30$ ; see Table 7). The descriptive statistics in Table 5 reveal that learners enjoyed the tour significantly more than expected. A significant effect of study was found on learners guide-related enjoyment ( $F(1,27) = 11.90, p < 0.05, \eta^2_p = .31$ ), where learners enjoyed interacting with the guide's in Study 2 significantly more than those in Study 1.

---Table 7 about here ---

Open-ended emotion-related responses to the two questions were similar across studies where approximately half of the participants volunteered a positive emotion-related statement (or absence of negative emotion statement).

## Discussion

The aim of this study was to evaluate the effectiveness of two mobile-based AR applications, used together, to foster positive emotions and meaningful learning in a laboratory (Study 1; lab-based, virtually-situated location) as well as an outdoor, location-based study (Study 2; outdoor, location-based). In summary, across studies: (1) most learners were able to independently identify one of the three high-level differences between the past and present versions of the Roddick Gates and most of the specific historical differences were discussed with some

scaffolding from the guide; (2) learners reported enjoying the learning session; and (3) learners frequently used the app-based features to help them make comparisons between the past and present-day location. These findings demonstrate that learners were, in line with our hypotheses, able to effectively and enjoyably learn about historical differences between past and present historical locations by contextualizing multi-media representations. Furthermore, the findings demonstrate that the two mobile, AR apps were effective both in and outside of the laboratory.

In comparing the results from Studies 1 and 2, findings revealed that learners were able to identify more differences outdoors (on-site) and required less scaffolding to identify differences (in particular, for Transportation and Gates) and fewer answers being provided by the tour guide. Findings also suggest that scaffolding prompts were effective in eliciting historical differences from learners, although they were more necessary in the lab-based than the outdoor study. These findings are in line with our hypothesis that fostering historical reasoning through the integration of inquiry-supportive elements (posing historical questions, using sources of information, and contextualization) would lead to effective learning.

An unexpected result was that few participants were able to identify that the porter's house no longer exists in the contemporary Roddick Gates location, despite reading a short text about it before beginning to identify historical differences. This finding suggests that learners may be better able to identify historical differences between visual/pictorial information or similar channels of information than different channels of information. A greater number of participants noticed the absence of the porter's house in the outdoor study, which could point to a greater attention to detail (and more effective historical reasoning) in authentic (in-person) vs. virtual environments. The other results pertaining to

learning outcomes supports this hypothesis, though further research into this question is necessary.

In looking at pre to post session self-reports of emotions, we found that participants enjoyed the tour (virtually or on-location) more than they anticipated, enjoyed learning about the Roddick Gates about as much as they anticipated, and enjoyed interacting with the guide slightly less than anticipated in Study 1 and significantly more in Study 2. Although mean levels of enjoyment were not significantly different between studies, learners reported more enjoyment and less boredom across post session questionnaires for the outdoor study (Study 2). This finding was consistent with our hypothesis that enjoyment would be higher in the outdoor setting compared to the lab setting (Study 1) given (1) that the real-world environment may reduce cognitive load (by reducing discrepancies between the multiple representations of the setting) and as well as heighten relevance and realism. These findings are consistent with both Mayer's (2005; 2009) multimedia theory and Pekrun's (2006) control-value theory of achievement emotions.

Given that all mean levels of enjoyment were above 3.0, and all post-session levels approached or were at 4.00, we can conclude that the two apps were effective in promoting positive emotions; further supported by corresponding low mean levels of boredom. The fact that participants enjoyed visiting the Roddick Gates significantly more than anticipated at the outset of the lab-based study suggests that they enjoyed having their tour augmented by the apps and is supported by students' open ended responses to the two questions that followed the emotion questionnaire (*What do you like (and not like) about the McCord Urban Museum iPhone application?*). Specifically, 86% of the emotion-related responses were positively-valenced (or non-negatively-valenced) across studies. Although originally designed to provide usability feedback for future iterations of the app, the responses to these questions provide supplemental and converging

support for our self-report measures regarding participants' enjoyment from participating in the experiment about historical reasoning with a mobile AR app. Overall, emotion-related results are consistent with our hypothesis that the apps and their use in Studies 1 and 2 would promote positive emotions given the inclusion of certain app and experimental design features intended to enhance learners' perceptions of control and value (e.g., user-directed, multi-media, high realism, personally relevant).

Eye-tracking results indicated that learners used the apps to make frequent historical comparisons, a learning behavior which is a component of historical inquiry (van Drie & van Boxtel, 2008).

A limitation of the comparison between studies was that the human guides changed. In Study 1 two postdoctoral fellows, both men in their late twenties, served as the guide, whereas two female undergraduate research assistants in their early twenties served as the guides in Study 2. It is therefore possible that the significant difference in enjoyment attributed to the guide came from learners interacting with a guide of the same gender and closer to their own age (in most learners cases), as other studies have found with virtual pedagogical agents (artificial intelligence-driven avatars in tutoring systems; Arroyo, Burleson, Tai, Muldner, & Woolf, 2013). A second limitation was that the battery for the guide's phone (using the MetaGuide app) died during the sessions for two of the outdoor participants. This required the guide to show the learners pictures of the evolution of the Roddick Gates over time and the now-demolished Porter house on back-up print-outs rather than on the screen of the phone. This technical problem did not appear to have a strong effect on the learners' enjoyment of the study and did not hinder their identification of historical differences (the guide used a script to provide the prompts and feedback found in Table 2). A larger sample would further enhance the generalizability of the findings and conclusions of this study,

as would measurements of learning that lean themselves to inferential rather than descriptive analyses. Finally, while our research question that concerned the eyetracking data only called for areas of interest which we were able to accurately and efficiently capture using human coding, more sophisticated eyetracking analyses such as investigations of saccades could have been examined

In conclusion, these results suggest that the two apps can be effectively used synchronously to support history learning and promote positive emotions for lab-based (virtual) as well as in person (location-specific) tours. Moreover, this study's results provide an example of theoretically and empirically driven design features (e.g., historical reasoning through inquiry, user-directed, multi-media, high realism, personal relevance) being successfully integrated into an AR mobile app learning environment to lead to positive learning outcomes and emotions as well as adaptive learning behaviors.

The results from this study will also be used to automate the guide's function using artificial intelligence (AI). Programming AI into an app is costly and it was therefore determined that it would be best to evaluate the prompts the Guide provided (as this manuscript has outlined) using a human experimenter before doing so. For this next step, a decision tree will be designed and programmed into the app using Figure 4 and Tables 2 and 3, which informs guide-related messages in response to answers entered by participants. Additionally, an embodied pedagogical agent can be integrated into the interface of the MTL Urban Museum app that would be male or female, depending on the participants' self-reported gender, and have dialogue reflecting an early-twenties age range. Replacing the human guide with a pedagogical agent and integrating the MetaGuide app features into the MTL Urban Museum app will make it scalable to both a larger audience (e.g., educational field and museum trips, educational

tours) as well as use additional historical locations the MTL Urban Museum app is programmed to provide information about.

Future development work on the MTL Urban Museum app will also target increasing the dynamic nature of the information, visuals, and prompts the app will provide by exploring additional technologies. Examples include using markers to automatically change historical content based on information such as learner movement or point-and-scan approaches developed for mobile phone cameras (Broll et al., 2008; Zhou et al., 2008). Moreover, while Study 1 illustrated that learners frequently compared historical information between the technologies provided it could reduce cognitive load to integrate virtual and real-world information into a single interface using a visual over-lay approach (Wu et al., 2013). In order to accomplish this objective, head-mounted displays or glasses could be used to project virtual information onto the learners line of sight. Alternatively, a mobile phone's built in camera could be used to scan markers and make virtual object's appear to be overlaid with the real world being viewed through the phone.

Additional future directions will include integrating the content of the tour (potentially a different site) with the curriculum of a high school or undergraduate history course (increasing the quantity of content in the process) and take the shape of an educational field trip where students download a new version of the MTL Urban Museum with the human guide's role (e.g., scaffolding; see Table 3) integrated and automated. As such, the virtual guide's role would be similar to that of a human tutor or museum tour guide. In adapting the study and the apps for a curriculum-tied field trip it is possible that the historical topic may represent a more dramatic and controversial topic such as war or prohibition. We hypothesize that selecting a less neutral and more dramatic and dynamic topic that is also tied to students' curriculum (and therefore grades) could enhance appraisals of value

and increase enjoyment. On the other hand, such topics could also contain material that students could be sensitive toward, and therefore would need to be chosen with caution and potentially necessitate debriefing and empathetic instructional strategies being interwoven into the field trip and virtual guide's pedagogical repertoire (for examples see Harley, Lajoie, Frasson, & Hall, 2015).

Understanding which features of the historical learning session elicited the emotions students felt is another area for future work. Although learners' emotions were examined in relation to different object foci (i.e., stimuli: the tour, learning, and the guide; see Table 5), and emotion-related statements were identified from open-ended responses about the application, a better understanding of the role the apps themselves played in fostering emotions could be obtained, and provide additional and more complete affective information that could help leverage strengths and address weaknesses. In order to accomplish this objective questionnaires designed to ask learners about stimuli-specific emotions will be created based on work by Harley and colleagues (Harley, et al., in press). Finally, a longer experimental session may help reveal whether the positive emotions experienced while learners interacted with the apps is sustained beyond the relatively short experimental session Studies 1 and 2 examined.

This study has provided a valuable empirical evaluation of the synchronous use of two theoretically driven AR mobile apps that both supports the continued development and use of AR and mobile technologies for history learning as well as evaluating said learning technologies in both lab and on-location environments. Moreover, this study provides evidence of the benefits of designing and evaluating AR mobile apps using theories of education, psychology, and multimedia. In particular, historical reasoning through inquiry was supported through (1) integrating multi-media (e.g., text, images) into mobile AR apps as sources of historical information; (2) posing historical questions (e.g.,



guide prompts) that necessitate the use of substantive and meta-concepts as well as implicit argumentation; and (3) situating learners in-person or virtually in the present day location (i.e., contextualization) to foster history learning. Additionally (4) allowing users to direct their attention freely to different historical cues and information (e.g., apps) and (5) selecting personally relevant historical locations that are visited by the learner either in-person or virtually also supported history learning. These design features also supported enjoyment, likely because of the appeals to value (personal relevance), control (autonomy regarding attention to different historical cues), and (6) use of appealing design features (e.g., attractive content, graphics, technology). Finally, (7) the straight forward, user-friendly design of the MTL Urban Museum app likely fostered comparisons between informational sources because by keeping learners cognitive load (at least regarding the app) low. These AR mobile app design features can be seen as empirically tested guidelines for building similar educational apps and stand to inform the field as they are the first of their kind.

## **Acknowledgements**

The research presented in this paper has been supported by a Joseph-Armand Bombardier Canada Graduate Scholarship for Doctoral research from the Social Sciences and Humanities Research Council (SSHRC) and a postdoctoral fellowship from the Fonds Québécois de recherche - Société et culture (FQRSC) awarded to the first author. This research has also been supported by funding from the Social Sciences and Humanities Research Council of Canada awarded to the fifth author. The authors would like to thank Laura Pipe, Emilia Gonzalez, Kexin Li, Mikaela Morton, and Hasagani Tissera for assisting in running participants, coding, and transcribing data. The authors would also like to thank Kevin Kee,

Reza Feyzi-Behnagh, Gregory Trevors, and Daniel Beaudin for their feedback on the paper.

## References

Arroyo, I., Burelson, W., Tai, M., Muldner, K., & Woolf, B. P. (2013) Gender differences in the use and benefit of advanced learning technologies for mathematics. *Journal of Educational Psychology, 105*, 957-969.

Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE Computer Graphics and Applications, 21*(6), 34-47.

Bondareva, D., Conati, C., Feyzi-Behnagh, R., Harley, J., Azevedo, R., & Bouchet, F. (2013). Inferring learning from gaze data during interaction with an environment to support self-regulated learning. In C. H. Lane, K. Yacef, J. Mostow, P. Pavik (Eds.), *Lecture Notes in Artificial Intelligence: Vol. 7926. Artificial Intelligence in Education* (pp. 229-238). Berlin, Heidelberg: Springer-Verlag.

Bressler, D. M. (2013). Museums: Gateways to Mobile Learning. In Z. Berge & L. Muilenburg (Eds.), *Handbook of Mobile Education* (pp. 224-234). Abingdon: Routledge.

Broll, W., Lindt, I., Herbst, I., Ohlenburg, J., Braun, A. K., & Wetzel, R. (2008). Toward next-gen mobile AR games. *IEEE Computer Graphics and Applications, 28*(4), 40-48.

Bronack, S. C. (2011). The role of immersive media in online education. *Journal of Continuing Higher Education, 59*(2), 113–117.

- Cocciolo, A. & Rabina, R. (2013). Does Place Affect User Engagement and Understanding? Mobile Learner Perceptions on the Streets of New York. *Journal of Documentation*, 69, 98-120.
- DeLucia, A., Francese, R., Passero, I., & Tortora, G. (2012). A collaborative augmented campus based on location-aware mobile technology. *International Journal of Distance Education Technologies*, 10, 55–73.
- D'Mello, S. K., Lehman, B., & Person, N. (2010). Monitoring affect states during effortful problem solving activities. *International Journal of Artificial Intelligence in Education*, 20(4), 361-389.
- Jaques, N., Conati, C., Harley, J. M., & Azevedo, R. (2014). Predicting affect from gaze behavior data during interactions with an intelligent tutoring system. In S. Trausan-Matu., K. Boyer., M. Crosby., K. Panourgia (Eds.), *Lecture Notes in Computer Science: Vol. 8474. Intelligent Tutoring Systems* (pp. 29-38). Switzerland: Springer.
- 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., Carter, C.K., Papaionnou, N., Bouchet, F., Azevedo, R., Landis, R. L., & Karabachian, L.R. (in press). Examining the predictive relationship between personality and emotion traits and students' agent-direct emotions: Towards emotionally-adaptive agent-based learning environments. *User Modeling and User-Adapted Interaction*.
- Harley, J. M., Lajoie, S. P., Frasson, C., & Hall, N.C. (2015). An integrated emotion-aware framework for intelligent tutoring systems. In C. Conati & N. Heffernan (Eds.), *Lectures Notes in Artificial Intelligence: Vol. 9112. Artificial Intelligence in Education* (pp. 620-624). Switzerland: Springer.
- Ioannidis, Y., Vayanou, M., Iatropoulou, K., Karvounis, M., Katifori, V., Kyriakidi, M., ... & Triantafyllidi, M. L. (2011). Preferences and attitudes for personalized information provision. *Data Engineering*, 34, 36-41.
- Katifori, A., Karvounis, M., Kourtis, V., Kyriakidi, M., Roussou, M., Tsangaris, M., Vayanou, M., Ioannidis, Y., Balet, O., Prados, T., Keil, J., Engelke, T., & Pujol, L. (2014). CHESS: Personalized Storytelling Experiences in Museums. In A. Mitchell (Ed.), *The Seventh International Conference*

on *Interactive Digital Storytelling*, LNCS 8832 (pp. 232–235). Springer International Publishing Switzerland.

Keil, J., Zollner, M., Becker, M., Wientapper, F., Engelke, T., & Wuest H. (2011). The House of Olbrich—An augmented reality tour through architectural history. *IEEE International Symposium on Mixed and Augmented Reality*, 15-18.

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 Augmented Reality in museums. In *Digital Heritage International Congress (DigitalHeritage)* (vol. 2, pp. 685-688). Marseille, France.

Klopfer, E. (2008). *Augmented learning: Research and design of mobile educational games*. Cambridge, MA: MIT Press.

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.

Martin, S., Diaz, G., Sancristobal, E., Gil, R., Castro, M., & Peire, J. (2011). New technology trends in education: Seven years of forecasts and convergence. *Computers & Education*, 57, 1893-1906.

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, 243–258.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31-48). New York: Cambridge University Press.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York: Cambridge University Press.

Mayer, R. E. (2014a). Incorporating motivation into multimedia learning. *Learning and Instruction*, 29, 171-173.

Mayer, R.E. (2014b). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 43-71). Cambridge, MA: Cambridge Press.

Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1994). Augmented reality: a class of displays on the reality–virtuality continuum. *Proceedings the SPIE: Telem manipulator and Telepresence Technologies*, 2351, 282–292.

Osborne, J.W. A. Overbay (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment Research Evaluation*, 9, 1–8.

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. M., Stupnisky, R. H., Perry, R. P. (2010). Boredom in achievement settings: Control-value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology*, 102, 531–549.

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, 34–48.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. (2002). Academic achievement emotions in students' self-regulated learning and achievement: A program of quantitative and qualitative research. *Educational Psychologist*, 37, 91–206.

Pekrun, R., Hall, N. C., Goetz, T., & Perry, R. (2014). Boredom and academic achievement: Testing a model of reciprocal causation. *Journal of Educational Psychology*, 106, 696–710.

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: Routledge.

Pujol, L., Katifori, A., Vayanou, M., Roussou, M., Karvounis, M., Kyriakidi, M., Eleftheratou, S., & Ioannidis, Y. (2013). From personalization to adaptivity - Creating immersive visits through

- interactive digital storytelling at the Acropolis Museum. *Workshop Proceedings of the 9<sup>th</sup> International Conference on Intelligent Environments*, pp. 541-554.
- Rabina, D., & Cocciolo, A. (2012). Uncovering lost histories through GeoStoryteller: A digital GeoHumanities project. *Digital Humanities 2012*, Hamburg, Germany.
- Rennick-Egglestone, S. J., Roussou, M., Brundell, P., Chaffardon, C., Kourtis, V., Koleva, B., & Benford, S. (2013). Indoors and outdoors: Designing mobile experiences for Cite de l'Espac. In *NODEM (Network of Design and Digital Heritage) 2013 Conference: Beyond Control – The Collaborative Museum and its Challenges* (pp. 89-97). Stockholm, Sweden.
- Rosenbaum, E., Klopfer, E., & Perry, J. (2007). On location learning: authentic applied science with networked augmented realities. *Journal of Science Education and Technology*, 16, 31–45.
- Rothfarb, R. (2011). *Mixing realities to connect people, places, and exhibits using mobile augmented-reality applications*. Paper presented at the Meeting of Museums and the Web, Philadelphia, PA.
- Rouet, J. F., Britt, M. A., Mason, R. A., & Perfetti, C. A. (1996). Using multiple sources of evidence to reason about history. *Journal of Educational Psychology*, 88(3), 478–493.
- Schreiber, W., Körber, A., Von Borries, B., Krammer, R., Leutner-Ramme, S., Mebus, S., Schöner, A., & Ziegler, B. (2006). Historisches Denken. Ein Kompetenz-Strukturmodell. [Historical Thinking. A model. A model of competences] Ars una, Neuried, Germany.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson Education/Allyn and Bacon.
- Tallon, L. (2013) Mobile strategy in 2013: An analysis of the annual museums and mobile survey. Pocket-Proof. Retrieved from: <http://www.museums-mobile.org/survey>
- van Drie, J., & van Boxtel, C. (2008). Historical reasoning: Towards a framework for analyzing students' reasoning about the past. *Educational Psychology Review*, 20, 87-110.
- van Gog, T. (2014). The signaling (or cueing) principle in multimedia learning. In R. E. Mayer (Ed.). *The Cambridge Handbook of Multimedia Learning* (pp. 263-278). New York, NY: Cambridge University Press.

Vayanou, M., Karvounis, M., Kyriakidi, M., Katifori, A., Manola, N., Roussou, M., & Ioannidis, Y. (2012). Towards Personalized Storytelling for Museum Visits. *The 6th International Workshop on Personalized Access, Profile Management, and Context Awareness in Databases (PersDB 2012)*. Istanbul, Turkey.

Vayanou, M., et al. (2014). Authoring personalized interactive museum stories. In A. Mitchell (Ed.), *The Seventh International Conference on Interactive Digital Storytelling, LNCS 8832* (pp. 37–48). Springer International Publishing Switzerland.

Watt, J. D., & Vodanovich, S. J. (1999). Boredom proneness and psycho- social development. *Journal of Psychology: Interdisciplinary and Applied*, 133, 303–314.

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

Zhou, F., Duh, H. B. L., & Billingham, M. (2008). Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. In the *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality* (pp. 193-202). IEEE Computer Society.

Table 1. *Summary of studies*

Feature	Study 1 (lab-based study)	Study 2 (outdoor study)
Location	Virtual view of the Roddick Gates (using Google Earth App displayed on SMART board)	Actual view of the Roddick Gates (in person)
Mobile Apps (and hardware)	<u>MetaGuide</u> (on Android; used by human guide)  MTL Urban Museum (on iPhone; used by learner)	<u>MetaGuide</u> (on Android; used by human guide)  MTL Urban Museum (iPhone; used by learner)
Self-report measures	Pre and post academic achievement emotions filled out on desktop	Pre and post academic achievement emotions filled out on iPhone
Other Equipment	Portable SMI eye tracking glasses	N.A.
Research Questions	(1) Were learners able to identify historical differences between the past and present? (2) Did learners report enjoying their learning session? (3) Did learners use the app-based features designed to help them make comparisons and provide additional context?	(1) Were learners able to identify historical differences between the past and present? (2) Did learners report enjoying their learning session?



Table 2. *Dialogue coding scheme: Identification of historical differences*

Level of Difference	Code	Definition	Example
Superordinate	Independently (IND)	The learner was able to independently identify a general historical difference.	"...I mean the gate itself." (PN12)
	Told (TOLD)	The guide provided the learner with a general (superordinate) historical difference.	"OK, so based on these pictures can you try and explain how transportation today has changed from transportation in the 1800s?" (PN02)
Subordinate	Independently (IND)	The learner was able to independently identify a specific historical difference.	"Yeah, now it was small before and now it is system one and the pillars, I guess the style, stylish pillars, its metallic or Doric stone pillar." (PN08)
	Told (TOLD)	The guide provided the learner with a specific (subordinate) historical difference.	"In terms of transportation you see the horse carriages and stuff, like obviously it's not commonly used" (PN03)
	With Help (SCAFF)	Learner was able to identify a specific difference after being given a prompt to elaborate or a direction to move in (high-level; superordinate).	"[Guide] So what about the Rddick Gates themselves...?" [Learner] "Well, they are built differently... and they are like columns today." (PN06)

Table 3. *Dialogue coding scheme: Guide prompts and feedback*

Type of Prompt	Definition	Example
Open-ended (OE)	When the guide asks the learner, generally, if they can identify any differences between the two versions of the Roddick Gates.	“Can you notice any differences between the past and the present?” (PN01)
Elaboration (ELAB)	When the guide asks the learner to expand upon a difference they have pointed out between the two versions of the Roddick Gates.	“Any other changes you can see in terms of transportations specifically?” (PN03)
Provide Context (CNT)	When the guide provides the learner with additional information about historical differences; oftentimes by showing them images and text from the Guide’s version of the MetaGuide App.	“...this is what the porter house used to look like. The porter house was constructed by the principle at that time, Sir John William Dawson” (PN04)
Clarification (CLR)	When the guide answers a procedural question the learner poses.	“Yeah, you can go back if you’d like to.” (PN02)
Agreement (AG)	When the guide notes their recognition of and agreement with something the learner has said.	“Yeah, no parking meters in that time.” (PN01)
Provide Answer (ANS_SUP)	When the guide provides the learner with a high-level (superordinate) historical differences.	“Can you describe how transportation has changed?” (PN04)
Provide Answer (ANS-SUB)	When the guide provides the learner with a specific / low-level (subordinate) historical differences.	“Still had the actual wooden fence...” (PN02)

Table 4. *Learners’ identification of historical differences by guide support for studies 1 and 2.*

Historical difference	Identified	Study
-----------------------	------------	-------

		1 (N = 13)		2 (N = 17)	
		Total	Participants (%)	Total	Participants (%)
Transportation	Independently	4	4 (0.31)	12	7 (0.41)
	With scaffolding	16	13 (1.00)	18	13 (0.76)
	By guide	4	4 (0.31)	1	1 (0.06)
	All	24	13 (1.00)	31	17 (1.00)
Buildings	Independently	3	3 (0.23)	10	6 (0.35)
	With scaffolding	8	7 (0.54)	9	9 (0.53)
	By guide	12	12 (0.92)	13	13 (0.76)
	All	23	13 (1.00)	32	17 (1.00)
Gates	Independently	16	7 (0.54)	24	14 (0.82)
	With scaffolding	16	10 (0.77)	6	4 (0.24)
	By guide	6	5 (0.38)	7	5 (0.29)
	All	38	13 (1.00)	37	13 (0.76)
Across all	Independently	23	9 (0.69)	46	15 (0.88)
	With scaffolding	40	13 (1.00)	33	16 (0.94)
	By guide	22	12 (0.92)	21	13 (0.77)
	All	85	13 (1.00)	100	17 (1.00)

*Note.* The values in the Total columns correspond to the number of subordinate historical differences identified across all participant tours. The values in the Participants (%) column correspond to the number of participants whom identified one or more historical difference independently, with scaffolding from the guide, or by having the guide identify the historical difference for them.

Table 5. Learners' self-reported enjoyment and boredom for studies 1 and 2

Test	Emotion	Object focus	Study 1 Mean (SD)	Study 2 Mean (SD)	Item
PreTest					
	Enjoyment	Tour	3.15(1.34)	2.94(1.12)	I get excited about going to visit the Roddick Gates tour location
		Learning	3.62(0.87)	3.88(1.02)	I look forward to learning about the history of the Roddick Gates
		Guide	3.54(0.78)	4.00(0.89)	I look forward to talking with the tour guide
PostTest					
	Enjoyment	Tour	3.69(0.95)	3.81(1.04)	After visiting the Roddick Gates I start looking forward to the next location
		Learning	3.69(1.10)	3.81(1.17)	Reflecting on my progress in learning about the Roddick Gates makes me happy
		Guide	3.23(1.01)	4.19(0.75)	I enjoy talking with the tour guide
	Boredom	Tour	2.08(1.19)	1.63(0.81)	Because I get bored at the Roddick Gates tour location my mind begins to wander
		Learning	1.85(0.80)	1.50(0.82)	Learning about the history of the Roddick Gates bores me
		Guide	1.69(0.75)	1.56(0.63)	Talking with the tour guide bores me

*Note.* Responses were based on a 5-point Likert response scale where learners were asked to rate their response to the items where 1 corresponded to "strongly disagree" and 5 to "strongly agree."

Table 6. Learners gaze transitions between Google earth and history apps

PN	Total time	# Transition	Transition per min
1	4.16	64	25.00
2	6.53	43	10.41
3	5.06	44	14.40

4	4.45	37	13.00
5	6.50	61	14.88
6	4.17	42	16.34
7	3.45	31	13.80
8	3.4	37	20.10
9	3.7	22	12.00
10	2.19	51	37.00
11	4.22	73	27.90
12	3.28	61	29.32
13	4.17	61	23.74
Avg.	4.34 (1 min 34)	48.23 (14.95)	19.84 (8.12)

Note. Time reported in minutes and seconds. Time is the period from when the guide asks the participant to note differences between the contemporary and 1800 version of the gates and the conclusion of their learning session. # Transitions refer to the number of times a participant switched their gaze between Google Earth and the mobile applications (learners' and the guide's). Moving from the Google Earth program displayed on the SMART board to the McCord MTL Urban Museum app displayed on the learners' phone would constitute one transition.

Table 7. Summary of repeated measures and univariate ANOVAs for enjoyment and boredom

Emotion	Object-foci	Effect	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	$\eta^2_p$
Enjoyment	Tour	Within subjects	1, 27	7.12	11.73	0.02*	.30
		Interaction	1, 27	0.41	0.67	0.422	.02
		Between subjects	1, 27	0.03	0.02	0.90	.00
	Learning	Within subjects	1, 27	0.00	0.00	0.97	.00
		Interaction	1, 27	0.70	0.12	0.75	.00
		Between subjects	1, 27	0.52	0.35	0.56	.01
	Guide	Within subjects	1, 27	0.05	0.06	0.81	.00
		Interaction	1, 27	0.88	1.00	0.33	0.04
		Between subjects	1, 27	7.21	11.90	0.00**	0.31
Boredom	Tour	Between subjects	1, 27	1.47	1.48	0.23	0.05
	Learning	Between subjects	1, 27	0.86	1.31	0.26	0.05
	Guide	Between subjects	1, 27	0.12	0.26	0.62	0.01

Note. \*  $p < 0.05$ , \*\* $p < .01$



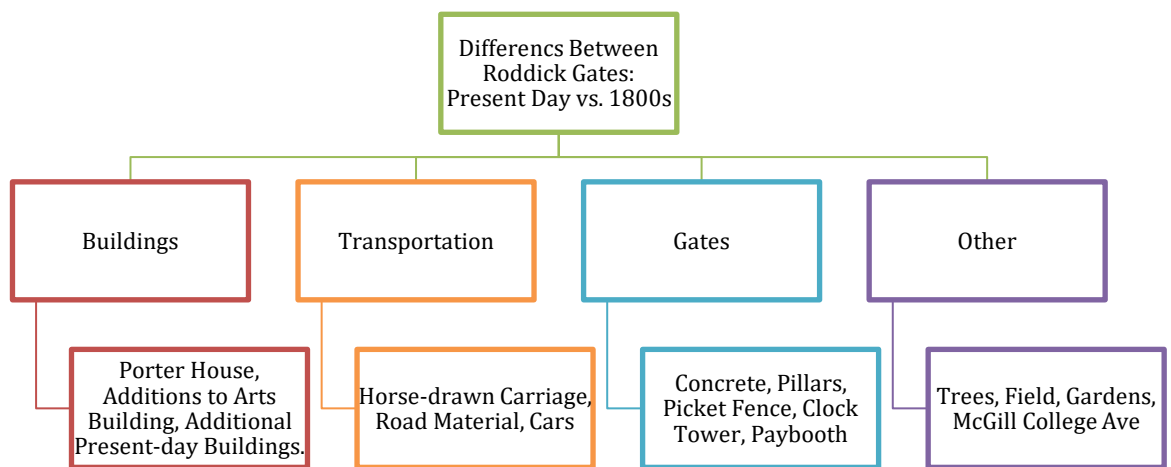
*Figure 1.* Picture of the MTL Museum App interface at the Roddick Gates location (text hidden).



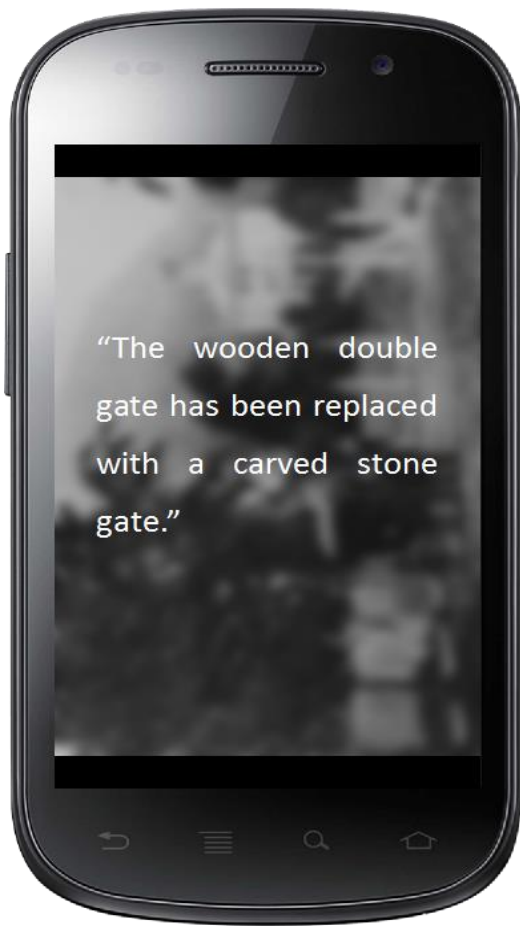
*Figure 2.* Screenshot from portable SMI eye tracking glasses (SMI ETG 2w; 30Hz) of learner using the MTL Urban Museum app and Google Earth to identify historical differences between the Roddick Gates in the past (1800s) and the present. The eye tracking video provided second-to-second information on what participants were looking at. The orange dot corresponds to what the participant is currently looking at during the video replay.



*Figure 3.* Picture of the main location-specific MetaGuide interface. Historical differences are outlined in red to remind the guide of the learning objectives for the learning session.



*Figure 4.* Three broad categories of historical differences between the present day Roddick Gates and the same gates during the 1800s. The lowermost boxes represent specific historical differences between the past and present day Roddick Gates for each of the three superordinate categories. The “Other” category represents differences not considered relevant to the learning objectives of the historical inquiry app interaction.



*Figure 5.* Picture of MetaGuide providing a prompt to the guide.