Fostering Positive Emotions and History Knowledge with Location-based Augmented Reality and Tour-guide Prompts

Jason M. Harley, Ph.D. University of Alberta

*Corresponding author

Assistant Professor, University of Alberta, Educational Psychology

Email: jharley1@ualberta.ca; *Phone:* (780) 492-9170

Mail: 6-102 Education North, Dept. of Educational Psychology, University of Alberta, Edmonton, Alberta, T6G 2G5, Canada

Susanne P. Lajoie, Ph.D. McGill University
Full Professor and Tier 1 Canada Research Chair, McGill University, Department of Educational and Counselling Psychology, 3700 McTavish Street 614, Montréal, QC, CAN, H3A 1Y2

Email: susanne.lajoie@mcgill.ca

Tara Tressel, MA McGill University
Ph.D. candidate, McGill University, Department of Educational and Counselling Psychology, 3700 McTavish Street 614, Montréal, QC, CAN, H3A 1Y2

Email: tara.tressel@mail.mcgill.ca

Amanda Jarrell, MA McGill University
Ph.D. candidate, McGill University, Department of Educational and Counselling Psychology, 3700 McTavish Street 614, Montréal, QC, CAN, H3A 1Y2

Email: amanda.jarrell@mail.mcgill.ca

Key words: augmented reality; mobile app; emotion; history; knowledge.

Acknowledgements and Funding: This research was supported by funding from the Social Sciences and Humanities Research Council of Canada (SSHRC).

Highlights

1. Significantly higher levels of enjoyment and curiosity from learning about history than from interacting with the app.
2. Learning outcomes significantly higher in the extended protocol than the previously developed one (high scores in both).
3. Significantly higher levels of task value reported after the guided tour compared to learners' pre-guided-tour responses.
4. Significantly higher levels of enjoyment from learning about the Arts Building than history learning in formal settings.
5. Findings paint a coherent and optimistic picture regarding the use of mobile AR apps for teaching history.

Fostering Positive Emotions and History Knowledge with Location-based

Augmented Reality and Tour-guide Prompts

Abstract

There is a lack of theory-driven empirical research that evaluates outcomes of location-based augmented reality (AR) applications with the purpose of improving instructional design and use guidelines. The primary aim of this study was to compare the effectiveness of two historical reasoning guide protocols, one based on prior research by Harley and colleagues (2016; the other an extension) while learners used a mobile AR app to learn about history. Learners reported significantly higher levels of enjoyment and curiosity from learning about history than using the app itself, though mean levels were high for both—in contrast to negative emotions. Results suggest that the new and extended historical reasoning guide protocol succeeded in fostering higher levels of knowledge than the former. Findings also revealed that learners reported significantly higher levels of task value after the guided tour compared to their pre-guided-tour responses. Implications and future directions are discussed.

Keywords: augmented reality; mobile app; emotion; history; knowledge

Fostering Positive Emotions and History Knowledge with Location-based Augmented Reality and Tour-guide Prompts

1. Introduction

Research on the effectiveness of augmented reality (AR) systems is in an early stage, and there is currently a paucity of empirical research with which to inform and guide the design and implementation of these systems, especially for educational purposes (Bacca, Baldiris, Fabregat, Graf, & Kinshuk, 2014; Poitras, Harley, Compeasu, Kee, & Lajoie, 2016; Wu, Lee, Chang, & Liang, 2013). In a recent review, Wu and colleagues (2013) summarized several features and affordances of AR for education: (1) learning content in 3D perspectives, (2) ubiquitous, collaborative and situated learning, (3) learners' sense of presence, immediacy, and immersion, (4) visualizing the invisible, and (5) bridging formal and informal learning. While these features and affordances are not unique to AR, they do paint an enticing portrait of the potential that AR holds for enhancing learning and students' enjoyment of it.

But what is AR? According to Wu and colleagues (2013), AR is best understood from a broad view as an approach used to add contextual, digital information to one's natural environment, thus *augmenting* one's experience of it. The broad view advocated by Wu and colleagues and others (e.g., Klopfer & Squire, 2008) focuses on the applications of AR rather than the specific technology it relies upon, such as head-mounted displays. A broad view of AR does not necessitate lumping all types of AR together, however. For example, mobile, location-aware AR (used in this study) utilizes global positioning system (GPS) technology to track learners as they physically move throughout real world locations and can augment the information they are presented with in response to location (e.g., changes to text or image-based historical information based on a change in location; see Zhou, Duh, & Billinghurst, 2008). Location-based mobile AR can promote learning, for example, about history through the delivery of digital media (i.e., archival sound, video, graphics, and texts) that is dynamically linked to

specific exhibits and artefacts 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 learners are engaging in with an AR-supported device.

Studies of location-based mobile AR in education have tended to rely primarily on self-reports of usability and preference, qualitative field observations, post-tour interviews, and system generated log-file data to evaluate their effectiveness (Bacca et al., 2014; Ioannidis et al., 2011; Harley, Poitras, Jarrell, Duffy, & Lajoie, 2016a; Huizenga, Admiraal, Akkerman, & Dam, 2009; Katifori et al., 2014; Keil et al., 2013; Rabina & Cocciolo, 2012; Vayanou et al., 2012). While many recent studies have formally measured post-mobile AR app interaction knowledge and found evidence that students learn better with mobile AR apps than low (e.g., digital book) or no-tech control conditions (Chang, Hou, Pan, Sung, Chang, 2015; Chiang, Yang, & Hwang, 2014; Efstathiou, Kyza, & Georgiou, 2018; Li, Zhang, Sundar, & Duh, 2013; Yoon, Anderson, Lin, & Elinich, 2017), much work remains to be done in this emerging area of research.

Indeed, beyond knowing whether mobile AR apps provide additive educational value to no or low-tech alternatives, research must also examine the pedagogical approaches best suited to leveraging the affordances of these technologies. This line of inquiry would benefit from carefully designed studies that compare pedagogical approaches, such as different instructional scaffolds. Such studies would provide evidence-based recommendations for how to optimize learning with AR. However, few studies have tested instructional or pedagogical adaptations or additions to mobile AR apps (Harley et al., 2016a), despite findings that how mobile AR apps are used matters for learning (Chang et al., 2015; Hwang, Wu, Chen, & Tu, 2016; Ibáñez, Di-Serio,

Villarán-Molina, & Delgado-Kloos, 2016). For example, Hwang and colleagues (2016) developed a competitive gaming approach to support AR-based learning activities conducted in a real-world field trip and found that elementary students assigned to the gaming approach outperformed students in the non-gaming approach. Chang and colleagues (2015) integrated AR with concept maps to form a concept-mapped AR (CMAR) scaffold and found that fifth-grade students assigned to the CMAR condition performed significantly better than those in the standard AR group. Ibáñez and colleagues (2016) developed a version of a mobile AR app for ninth grade students with task suggestions (based on their individual pre-test scores) and found that students using this version of the AR app outperformed those with the standard version of the AR app. Finally, Harley and colleagues (2016a) examined university students' ability to identify historical differences while using a mobile AR app with prompts from a human guide in an outdoor, on-location setting compared to an indoor setting where the historical location was virtually represented on a SmartBoard using Google Earth. They found that both settings were comparable, though descriptive statistics revealed preliminary, directional evidence that outdoor learners were able to identify more historical differences, required less scaffolding from the guide to do so, and enjoyed the tour more than those who completed the tour in an indoor, virtual setting.

The latter study also addresses another under-examined area of research with mobile AR app: learners' emotions. While studies have found motivational (e.g., Furió, Juan, Seguí, & Vivó, 2015; Chang, Hsu, & Wu, 2016), attitudinal (Hwang, Wu, Chen, & Tu, 2016), and other psychological benefits (e.g., "sense of place"; Chang, et al., 2015; Chen et al., 2016) from using mobile AR apps to learn, emotions have not been a focus of research—despite the critical implications these states have for learning and their

increasing prominence in educational research (Pekrun, 2006; Pekrun & Perry, 2014). This study contributes to addressing both of these gaps by evaluating two different human guide instructional scaffolding protocols (one from Harley et al., 2016a) designed to leverage university students' use of a mobile AR app and foster positive emotions, knowledge outcomes, and appraisals of task value.

1.2. Theoretical Framework

Our past and current mobile AR history research (Harley et al., 2016a) is guided by the control-value theory of achievement emotions (Pekrun, 2006; Pekrun & Perry; 2014) and a novel application of van Drie and van Boxtel's (2008) historical reasoning framework. We provide a brief overview of each theory below.

1.2.1. Emotions. For effective learning to take place, learners need to be in (and remain in) an emotional state that is amenable to concentrating, reasoning, and other learning-related processes (Jarrell, Harley, Lajoie, & Naismith, 2017; Pekrun, 2006; Pekrun & Perry, 2014). Emotions can enhance achievement by fostering motivation, and focusing attention and limited cognitive resources on achievement-related activities (e.g., when experiencing enjoyment of task), and promoting situationally-appropriate information processing and self-regulation strategies (Pekrun, Elliot, & Maier, 2009; Pekrun & Perry, 2014). Negative emotions such as boredom can, however, be maladaptive to achievement and undermine interest and intrinsic motivation (Pekrun & Perry, 2014) as well as consume cognitive resources needed for the achievement task (Meinhardt & Pekrun, 2003).

A central tenet of the control-value theory of achievement emotions (CVT; Pekrun, 2006; Pekrun & Perry, 2014) is the role appraisals of control and value play as proximal antecedents of emotions. Subjective control is defined as one's perceived ability to

effectively manage achievement activities and their outcomes, or more broadly, as one's beliefs concerning the causal influence they exert (agency) over their actions and outcomes (controllability), including the subjective likelihood of being able to obtain said outcome (probability). Subjective value is defined as the perceived importance of an activity or its outcome(s) to oneself (goal relevance), combined with the perception that an action or outcome is positive or negative in nature (goal congruence—event supports or hinders goal attainment).

Another important component of the CVT is the role that focusing one's attention (object focus) on academic achievement outcomes vs. an academic achievement activity has on the generation of emotions. Enjoyment from listening to an interesting lecture is an example of an activity emotion, whereas frustration from recalling one's low score on an exam perceived as unfair is an example of an outcome emotion. The CVT also draws a distinction between prospective, concurrent, and retrospective time frames for achievement emotions. When outcome foci are oriented toward the future, emotions are referred to as prospective emotions and when foci are oriented toward the past they are referred to as retrospective emotions. Concurrent emotions include emotions aroused from an activity one is currently undertaking. Together, appraisals and object foci influence and constrain emotional responses. For instance, if a learner feels highly in control of a task and also highly values a task, they are expected to experience (concurrent) enjoyment of the activity, experience hope or joy from looking forward to it, and/or take pleasure in recalling it.

These assumptions informed our measurement of emotions and design of the guided tour with the mobile AR app. For example, the MTL Urban Museum on the McCord Museum app provides user-directed navigation (e.g., choice in what they pay

attention to, for how long, and in what order), which can enhance learners' perceptions of control by supporting autonomy and self-directed inquiry. In terms of enhancing value, the interaction with the guide and McCord Museum app provides historical information about a real-world setting that is 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. Fostering positive emotions is consistent with recent extensions by Plass and Kaplan (2015) to Mayer's multimedia learning theory (2005), which posit that incorporating emotionally and motivationally appealing design features (e.g., attractive content, graphics) can help to increase cognitive engagement and retain learners' attention.

1.2.2. Historical reasoning. We define and operationalize historical reasoning as an activity during which learners acquire knowledge of the past and use it to interpret phenomena from the past and present (van Drie & van Boxtel, 2008). One way learners engage in historical reasoning is by comparing or explaining historical phenomena. In this regard, learners adopt an active role in building knowledge and an understanding 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. Historical reasoning has been examined in a variety of contexts, including schools and classrooms (Groot-Reuvekamp, Ros, & van Boxtel, 2018; Huijgen, van Boxtel, van de Grift, & Holthuis, 2017; Stoel, & van Drie, va, Boxtel, 2017), and museums (Marcus, Stoddard, & Woodward, 2017; van Boxtel, Grever, & Klein, 2016), to guide and enrich learners' development of historical knowledge. Research has found

evidence that learners who engaged in instructional interventions that supported historical reasoning outperformed learners who participated in more traditional approaches to learning history (Groot-Reuvekamp, et al., 2018; Stoel et al., 2017). Research on historical reasoning has, however, had little interaction with educational technologies, particularly mobile AR (van Drie, van Boxtel, & van der Linden, 2006; Marcus, Stoddard, & Woodward, 2017; van Boxtel, Grever, & Klein, 2016). Moreover, research on historical reasoning has focused its limited examination of emotions through qualitative analyses and the lens of empathy rather than examining learning-related (e.g., achievement) emotions (Efstathiou, Kyza, & Georgiou, 2018; Marcus, et al., 2017; van Boxtel, et al., 2016). Given the importance of emotions to supporting effective learning (Pekrun & Perry, 2014), this is an important gap to address.

This article therefore extends research with historical reasoning to an under-examined but promising (Efstathiou, et al., 2018) technological context while answering researchers' calls to attend to the emotions generated by museum visits and history education (Watson, 2015). Investigating emotions during museum visits is important because different learning environments have different characteristics, affordances, and constraints that can influence emotions as distal antecedents (Pekrun & Perry, 2014). For example, museums and tours, including tours with mobile apps, are typically less formal and more open-ended than traditional history learning environments (e.g., classrooms); characteristics that may have implications for the generalizability of findings on emotions and history education. Indeed, according to the control-value theory of achievement emotions, achievement emotions should be organized in both domain and task-specific ways (Pekrun & Perry, 2014). In contributing to addressing these gaps in the literature,

our work incorporates all six of the core components of the historical reasoning framework into the design and learning activity with the mobile AR app.

Our prior work utilized a human guide to pose historical comparison questions that learners used to engage in historical reasoning (Harley et al., 2016a). Learners used multimedia information, including images and text about a historic location to help contextualize (i.e., spatially and socially situate) the contemporary version of the location with the past. Argumentation took place when learners made evidence-based claims about how they knew a change had taken place. Substantive concepts such as 'transportation' and meta-concepts such as 'change' also played a role in how learners identified and reflected upon changes.

The results of this first study revealed that the posing of historical comparison questions was both effective and necessary for learners to make all the key historical comparisons in both a virtual and on-site tour context while interacting with a mobile AR app. However, we wanted to extend this work to include prompts that directly incorporated all aspects of historical reasoning, especially argumentation which is foundational in ensuring that assertions and claims about the past are supported by rational arguments and evidence, not opinion or viewpoint; an important distinction in contemporary education.

We call this new protocol, the extended prompt and feedback (EPF) protocol which is described more fully in the methods section. The EPF protocol builds off of the success of what we refer to as the comparison prompt and feedback (CPF; Harley et al., 2016a) protocol and is primarily enhanced with questions that support argumentation by asking learners causal and explanatory questions related to a new historical location on McGill University's campus. Additionally, learners were encouraged to use and coordinate

information from multiple sources of historical information that correspond to sites that were connected (e.g., different sections of the building of interest) or in close proximity to the main historical location (e.g., a monument in front of the building) in order to answer questions and appreciate how McGill University's campus has changed. A greater number of substantive concepts, such as important people and places, are touched upon in the EPF guide protocol as well. Additionally, meta-concepts such as evidence, time, place and change are more thoroughly reflected upon and explored.

According to the cognitive process dimension outlined in Krathwohl's (2002) revision of Bloom's taxonomy of cognitive skills, the CPF protocol fostered relatively simple cognitive processes that best corresponded to 'recognition' and 'remembering'. The EPF protocol extended the complexity of cognitive processes required of the learner by encouraging them to engage in processes outlined in the higher-level 'understanding' level of the hierarchy, such as 'interpreting' and 'explaining'. Increasing the sophistication and breadth of learners' cognitive processing and historical reasoning stands to enhance the quality of history learning by increasing the extent to which learners are cognitively active while learning; a critical ingredient in meaningful learning (Mayer, 2002).

In sum, the CVT (Pekrun, 2006; Pekrun & Perry, 2014) informed our measurement of emotions and use of the mobile AR app as well as further motivated the need to expand research on emotions in both mobile AR and historical reasoning research. The historical reasoning framework (van Drie & van Boxtel, 2008) on the other hand, informed our design of the instructional prompts the human guide administered to help the learner use the mobile AR app effectively to increase their history knowledge. Research has found that both emotions and historical reasoning can help students

improve their knowledge, this study is the first to examine both and in the context of mobile AR.

1.3. The Current Study

In this study, we evaluated two historical reasoning guide protocols used in conjunction with a mobile AR app to foster positive learner emotions, task value, and increased knowledge about the history of their university by focusing on an iconic landmark, including how and why it has changed. We report on our evaluation of the guide protocols and mobile AR app interaction here by focusing on our analyses and discussion of self-reported academic achievement emotions, objective and subjective measures of knowledge, and self-reported appraisals of task value.

The primary aim of this study was to (1) compare the effectiveness of the EPF guide protocol to the CPF guide protocol while learners used the mobile AR app. We operationalize effectiveness in this study in terms of (a) emotional engagement, (b) objective and subjective knowledge outcomes, and (c) value of history learning. Consistent with our first study (Harley, et al., 2016a) and the idea that what a learner is attending to (object focus; e.g., activity vs. outcome; Pekrun, 2006) will influence the generation of emotions, we examined learners' emotions directed toward: (i) learning about the Arts Building, (ii) the mobile AR app they used, and (iii) the guide's prompts and feedback. In order to contextualize the results within the larger picture of learners' feelings and value directed toward history, we pursued two further, complimentary aims: (2) to examine learners' appraisals of the value of learning about the history of McGill University, including appraisal levels before and after their guided tour; and (3) to compare how learners felt about learning about history during the guided tour with the

mobile AR app compared to typical formal and informal history learning contexts. To address the three objectives we investigated four research questions:

1.3.1. (RQ1) Did learners report feeling emotionally engaged during the tour? (a) What kinds of emotions did learners report experiencing toward different focal points? (b) Did learners' discrete emotions significantly differ based on the focal point that self-reported emotions were directed toward? (c) Did discrete emotions significantly differ between guide protocol conditions? Based on the CVT (Pekrun, 2006; Pekrun & Perry, 2014) and prior research (Harley, et al., 2016a) we hypothesized that learners would report high levels of positive emotions and low levels of negative emotions. Specifically, we hypothesized that the app would promote higher levels of positive, activating emotions (enjoyment and curiosity) and lower levels of boredom and frustration because the features of the AR learning environment and guide tour were designed in a way that should enhance learners' appraisals of control and value (see section 1.2.1). We also anticipated that encouraging learners to engage in more and varied types of historical reasoning in the EPF condition would increase their interest and appraisals of value about it, thus increasing levels of positive emotions, compared to learners who only received comparison-focused historical questions in the CPF condition. In other words, we hypothesized that the EPF condition was more likely to trigger learner curiosity and enjoyment by challenging them with additional prompts to do more of the thinking about how their university grew into its current form; an exercise that could help them further appreciate and value the historical facts the tour covered. No empirical literature that the authors were aware of was available to further inform this hypothesis.

1.3.2. (RQ2) Did learners in the EPF condition have higher (a) objective (post-tour test score) and (b) subjective (perceived success) knowledge outcomes

than those in the CPF condition? Based on prior research (Harley et al., 2016a), we anticipated that the guided tour with both protocols would effectively support historical knowledge outcomes. Accordingly, we hypothesized that learners randomly assigned to either condition would, at a minimum, pass the post-tour knowledge test (e.g., score higher than 50%) and report reasonably high levels of perceived success.

Although we hypothesized that learners randomly assigned to either condition would perform reasonably well, we anticipated that those in the EPF condition would have higher objective and subjective knowledge outcomes (since the two are typically correlated; Hall, et al., 2006; Ruthig, et al., 2007), than those in the CPF condition. We believed that engagement in a tour with a guide protocol designed to provide deeper historical reasoning, particularly, with regard to argumentation (van Drie & van Boxtel, 2008); the latter corresponding to higher levels of knowledge and cognitive processing (Krathwohl's, 2002; Mayer, 2002) would account for this. This hypothesis is also in-line with previous research linking historical reasoning to improved history knowledge post-test scores (see 1.2.2.).

1.3.3. (RQ3) How did learners appraise task value? (a) Did task value differ between conditions? (b) Did learners report significantly different levels of task value before and post the guided tour? Our hypothesis about appraisals of value (a proximal antecedent of emotions) between conditions was consistent with that mentioned above: higher levels of value in the EPF than CPF condition (see 1.3.1). We further hypothesized that learners would report valuing learning about history both before and after the tour, but that appraisals of task value would be higher post-tour. We anticipated that the learners who would volunteer to participate may have been motivated, in part, to participate because of the opportunity to learn about McGill's history (as advertised in

the participant recruitment materials). We hypothesized that a significant difference between pre-and post guided-tour levels of task value would be observed because learners' appraisals of value may have relied on a certain level of speculation regarding the nature and quality of the learning they would engage in. Once the tour was over, we believed appraisals would be higher because they would be based on actual versus speculated value of content.

1.3.4. (RQ4) Did learners report higher levels of positive emotions and lower levels of negative emotions from learning about the Arts Building after the tour compared to the emotions they reported experiencing during typical (a) formal and/or (b) informal learning contexts?

The proposed comparison between state emotions evoked during the study and emotional tendencies toward learning about history in formal and informal learning situations (trait emotions) is not a pure one because these represent different emotional constructs; the latter comprising more stable states, consisting of a longer time frame, multiple events, and being more susceptible to memory biases and global heuristics (Frenzel, Pekrun, & Goetz, 2007; Goetz, Bieg, Ludtke, Pekrun, & Hall, 2013; Scollon, Kim-Prieto, & Diener, 2003). State emotions, on the other hand, are defined as momentarily occurring emotions that are more strongly influenced by situational variables (Eid et al., 1999). While results have found weak to moderate correlations between trait and state emotions (Steptoe, Gibson, Hamer, & Wardle, 2007) as well as belief-driven discrepancies between what students think they feel (trait emotion) and what they actually feel (state emotion; Bieg, 2017), understanding relationships between these two types of emotions is none-the-less valuable. Indeed, if trait emotions, however flawed, serve as the comparisons learners use to evaluate the emotional engagement and

value of an activity, they stand to motivate behavior, including approaching and persisting with such activities in the future (Pekrun, 2006; Pekrun & Perry, 2014). In other words: if typical emotional dispositions toward topics such as history are what come to mind when learners think about the subject, than these states stand to influence individuals' behaviour (e.g., whether to attend a museum exhibit, inform themselves about how history has shaped contemporary events, etc.), and perhaps more so than state emotions because these states may be dismissed as unrepresentative. We were therefore particularly interested in how learners' emotions during the interaction with the mobile AR app and guide protocols compared to how they typically felt (i.e., remembered feeling) about learning history.

The CVT guided our hypothesis about the emotions learners reported experiencing prior to the guided tour in formal and informal contexts and post-tour. We hypothesized that learners would report higher levels of positive emotions in informal contexts than formal ones because informal contexts are typically more self-selected (suggesting higher value) and involve more autonomy (suggesting higher control) than formal ones where choice is limited and appraisals of value are extrinsic (instrumental; task valued for grade) rather than intrinsic (the task itself is valued). We hypothesized that the level of positive emotions learners would report about history (Arts Building focus) post-tour would be similar to emotions in informal settings because of its personal relevance.

2.0. Method

2.1. Participants

Forty-five university students (78% female[1]) from a large North American university volunteered to participate. Participants were between 19 and 32 years old (M = 23; SD = 3.1) and enrolled in various programs (e.g., 25 from science or engineering; four from business-related fields; and 10 from bachelor of Arts programs, including one BA double history and political science major). Student GPAs ranged from 2.20 to 4.00 (M = 3.44; SD = .42) out of four. There was no preference for participants' academic background and therefore no rationale for controlling demographics. In order to participate, students had to (1) be enrolled as a full-time student at McGill University at the time of the study and (2) not have prior experience using the MTL Urban Museum tours on the McCord MuseumUrban Museum app. Participants were compensated with $5 per half hour for a potential total of $10/hour. Compensation was provided at the end of the session. Participants were recruited from either the university's online classified advertisement or a McGill University undergraduate student Facebook group. Eligible participants were scheduled on a first contact-first schedule basis[2]. Participants were randomly assigned to one of the two protocol conditions and completed their individual guided tours with the same human guide[3] and mobile AR app. In addition to random

[1] We were unable to further match participant gender to specific participant numbers (and pre and post-tour surveys) for 18 of the 45 participants due to data storage problems with the demographic questionnaire. The gender breakdown of the 27 participants was 21 female (12 EPF, 9 CPF) and 6 male (4 EPF, 2 CPF). While examining the potential effects of gender were beyond the scope of this paper, we examined potential gender differences across all study variables. We found that learners' responses and scores did not significantly differ by gender, with the exception of perceived success. To examine the possible effect of gender on our results, we controlled for gender in supplemental analyses and found that when controlling for gender, the reported findings (see section 3) did not change, suggesting that gender did not have a significant influence on our results.

[2] Not all eligible students who contacted us to express their interest followed-up with providing availability or showing up on the day of the study. Moreover, time, human and financial resources did not permit us to run all of the students who expressed interest.

[3] The human guide was also one of the guides from the previous study to build on their prior knowledge of the CPF protocol (in a new location). In order to help prevent any meaningful differences in their behaviour between conditions they received approximately five to ten hours of training with the first author which involved explanations of the protocol and several rounds of mock data collection to ensure they stayed on-script. Moreover, the first author reviewed the first several audio recordings from data collection to ensure compliance with the protocol and a comparable, professional demeanour across conditions.

assignment, independent sample t-tests were conducted to test for potential pre-tour group differences. Results indicated that groups were statistically equivalent on all pre-tour variables (see Appendix A).

2.2. Mobile AR app

This study examined the effectiveness of a guided tour of historical landmarks on McGill University campus using the MTL Urban Museum on the McCord Museum app to foster emotional engagement and historical knowledge. The app was developed by the McCord Museum and allows users to locate historical landmarks in the city of Montreal using GPS smart phone technology. The app allows users to view multimedia content (historical text and images) that illustrate how the city used to look in the past (e.g., 1800s) in comparison to the present day. Figure 1 provides four captions of the app's interface and features used in the study. Caption 1 shows the app's homepage. Caption 2 shows a GPS-enabled map of Montreal, including the user's present location (represented by a blue circle) relative to nearby locations (red pins). Caption 3 is a zoomed-in version of the location-rich map, and includes a label of a nearby pin the user clicked on that corresponded to McGill's Arts Building. Users could access the historical multimedia by clicking on the information icon (the circled "i") on the right side of the label. Caption 4 illustrates a sample of historical multimedia that informs users about the Arts Building in the mid 1800's. The view of the Arts Building in the historical picture in caption 4 matches the view and geographical perspective of the Arts Building users had of the contemporary Arts Building during the guided tour. This view allowed users to visually compare how the Arts Building had changed by looking at a picture taken from the same position they were standing in long ago. Some locations, such as the Arts Building, have multiple pins that allow users to examine the same historical location from different

vantage points, focus on different sections of it, or learn about neighbouring landmarks. Figure 2 shows a tour guide interacting with a learner while she uses the app to learn about and compare text and visual information about the historical Arts Building with the contemporary version.

2.3. Historical Reasoning Prompt and Feedback Protocols

A human tour guide provided both procedural and pedagogical support to learners as they completed the experimental protocol (described below) and used the mobile AR app to increase their knowledge about the history of McGill University. The tour guide was an undergraduate research assistant trained by the first author on both prompt and feedback protocols. She also served as a guide from the previous study of historical reasoning using the same app and the CPF protocol at a different historical location. The guide used the protocol to interact with a single learner at a time. Tours in both conditions began with the same open-ended introduction (see open-ended introduction question in Appendix B) following a tutorial of how to use the app. A procedural clarification was offered if required. Appendix B also summarizes the other protocol-specific (see "scheduled" prompts) and protocol-general (see "general") prompt and feedback messages that the guide used in the tour.

In the CPF condition, learners were only encouraged to identify each of the four core differences between the historical and contemporary Arts Building. One of the four differences was that the mid 1800s version of the Arts Building did not have the east (Dawson Hall) and west (Molson Hall) wings now present in the contemporary building. Once learners in the CPF condition identified this difference the guide provided them with additional context underlying the historical meaning of the difference. To continue the example: that McGill University has grown. The EPF condition differed from the

CPF condition with the provision of a historical reasoning hint (e.g., "Why might these [wings] have been added?") designed to help learners figure out the context and meaning of the historical difference themselves (see Appendix B). Once the difference was uncovered, the guide then provided the additional historical context and facts about the difference that were provided in the CPF condition. Before moving on to prompt the learner to identify another difference (or end the session if all other differences had been identified), the guide provided historical facts related to the historical differences and its significance. For example, that *the faculty consisted of one Principle and four professors in 1821 when McGill was first founded (but before it formally opened). As of fall 2014 there were 1,674 full time professors at McGill University.* See Appendix C for a summary of the above example in table form and additional details.

2.4. Measures

2.4.1. Emotions. Learners' emotions were measured prior to the start of the tour and immediately after. Prior to the start of the tour learners were asked to report how they typically felt (trait emotions) about learning history in (a) formal and (b) informal academic contexts. A modified version of the Achievement Emotion Questionnaire (AEQ; Pekrun et al. 2002) was used for the wording of the questionnaires and items[4]. The questionnaire used a five-point Likert scale where 1 corresponded to "strongly disagree" and 5 corresponded to "strongly agree". Examples of formal (*elementary, high school, CEGEP, college/university curriculum*) and informal (*visiting a museum, watching a documentary, looking up something historical on Wikipedia, etc*.) academic contexts were given. Prior to answering the questions participants were prompted to *please recall some typical situations in which you learned about history.* Participants responded to

[4] Dr. Pekrun was also consulted on the wording of the items (R. Pekrun, personal communication, October 25, 2015).

questions asking them about the (prospective) academic achievement emotions they experienced before learning about history in formal and informal contexts. In order to avoid overburdening participants with long surveys a single item was used to assess enjoyment (I look forward to learning about history), boredom (because I find history boring, I have no desire to learn about it), and frustration (I'm annoyed when I have to learn about history) in formal and informal academic contexts. Participants were asked to respond to the same items in formal and informal contexts. See Pekrun, Vogl, Muis, and Sinatra (2016) and Harley, Bouchet, Hussain, Azevedo, and Calvo (2015) for similar single-item questionnaires for assessing discrete emotions.

After the guided tour, learners were asked to report how they felt *during* the tour. A modified version of the AEQ (Pekrun et al. 2002) and a prior version of this questionnaire (Harley et al., 2016a) was used for the wording of the questionnaires to assess learners' retrospective activity emotions. The questionnaire used a five-point Likert scale where 1 corresponded to "strongly disagree" and 5 corresponded to "strongly agree". Learners were asked to report how they felt about (a) learning about the Arts Building, (b) the mobile AR app they used, and (c) the guide's prompts and feedback (described to participants as the guide's "questions"). The questionnaire included twelve items; one for each of four emotions (enjoyment, boredom, frustration, and curiosity) for each of the aspects of the guided tour that may have aroused an emotion. The following example item was used to measure frustration aroused from the mobile AR app: *Using the McCord MTL Urban Museum App annoyed me.* Discrete emotions were selected based on the prevalence of emotions in technology-rich learning environments in similar studies and theoretical considerations regarding capturing different types of emotions that

could be expected to be elicited from the guided tour (D'Mello et al., 2013; Harley et al., 2016a).

2.4.2. History knowledge. History knowledge was primarily assessed using a seven-item multiple-choice test administered immediately following the tour. Each question contained five foils and was designed to assess learners' knowledge of the history of McGill University covered in the tour in relation to the Arts Building location. Each item was designed to assess different historical information and concepts[5]. See Appendix D for questions and foils.

Prior history knowledge was assessed by two questions administered by the guide in both conditions: (1) "Before we begin I want you to tell me everything you know about the history of McGill University. Take as much time as you need." And: (2) "Now please tell me everything you know about Montreal and Canada in the 1800s." Learners' transcripts revealed low levels of prior knowledge, including few accurate facts and the majority of learners explicitly stating that they "didn't know much"[6]. In order to better compare learners' expressed prior knowledge with our post-tour test, one graduate student and one undergraduate research assistant coded all of the transcripts for answers to the seven multiple-choice questions with an agreement of 100%. Coding supported our holistic assessment of learners' low prior knowledge where transcripts contained fewer than one of the seven answers in the post-tour knowledge test, on average (M = .10; SD = .09). Given the differences in how knowledge was assessed pre and post-tour, however,

[5] Given that each item was designed to assess different historical information and concepts and was not organized around subscales, we did not necessarily expect nor observe high internal reliability: Cronbach's Alpha indicated that internal reliability of the seven items was low (α = .25). This indicates that the post-tour test was not always measuring the same type of knowledge of history, as we suspected. See 4.1.

[6] 29 of the 39 (74%) students that we had transcript data from explicitly reported that they didn't know much in response to the prior knowledge prompts. E.g., "No clue" (PN 35), "No…*laughs* cause I'm not Canadian…" (PN31), "I really have no idea." (PN 19).

we opted not to use a proportional learning gain, as is often used when sufficient learning material is available to construct counter-balanced pre and post-educational intervention multiple-choice tests. Additionally, audio data from six participants was not available, so using this variable to construct a proportional learning gain or adding it as a covariate (despite no significant differences in prior knowledge being detected) would have reduced our sample size from 45 to 39 participants and negatively impacted our statistical power. Therefore, the prior knowledge score was only added as a covariate to supplement statistically significant effects related to RQ2.

2.4.3. Perceived success. In addition to measuring learners' post-tour knowledge of McGill's history, we were interested in how successful they *thought* they were. Previous studies have found high correlations between perceived success and actual achievement making this a good supplemental (subjective) measure of learning (e.g., r = .78, Hall, et al., 2006; r = .70, Ruthig, et al., 2007; e.g., r = .67). In order to measure perceived success we modified the question and item used in Daniels and colleagues (2008) to suit our study: ''How successful do you feel you were in learning about the history of McGill University from the guided tour you just completed?'' (1 = ''very unsuccessful'' to 10 = ''very successful'').

2.4.4. Task value. Task value refers to the students' evaluation of how interesting, how important, and how useful the task is perceived to be. Appraisals of value are also instrumental in influencing the arousal of emotions. In order to measure task value we adapted five items from the Motivated Strategies for Learning Questionnaire (MSLQ; Pintrich, 1991) for this study. A seven-point Likert scale was used to assess learners' task value where a ranking of 1 corresponded to "not at all true of me" and a ranking of seven corresponded to "very true of me". An example item from the

pre-tour survey was the following: *understanding the history of McGill is very important to me.* An example item from the post-tour survey was: "*I am very interested in the content on McGill history that was presented in this study*." Cronbach's Alpha indicated that scale reliability was high for the task value questionnaire administered before (α = .88) and after (α = .85) the guided tour.

2.5. Materials

Learners used an iPhone 5 to interact with the MTL Urban Museum on the McCord Museum app. Apple ear buds were used to record the learner-guide interactions, learners' prior knowledge activation, and historical reasoning from a think aloud protocol. Google Surveys was used to administer the questionnaires and Survey Monkey was used to administer the eligibility and demographics form.

2.6. Experimental Procedure

Experimental sessions took approximately half an hour to complete. The session began with the guide welcoming the participant at the historical location. Next, the learner read and signed the consent form before filling out the pre-tour questionnaires (emotions about history in formal and informal contexts and task value of history learning). After completing the questionnaires audio recording equipment was set up and started recording. Prior knowledge was activated by asking participants to tell the guide everything they knew about the history of McGill University. Participants were told to take as much time as needed. Once finished they were then ask to tell the guide everything they knew about Montreal and Canada in the 1800s. After activating their prior knowledge participants received a tutorial on how to use the app. Participants then virtually navigated to the Arts Building and its surroundings using the GPS features of the app so that they could see their own position relative to the Arts Building and pins

corresponding to historical sites within and in its immediate proximity. The tour started with an open-ended question and ended once all pertinent historical locations and relevant contextual differences were identified and discussed. Immediately following the completion of the tour participants completed the post-tour measures, including the retrospective emotion questionnaire, task value, perceived success, and post-tour test, in that order. Participants were then paid and debriefed.

2.7. Data Analyses

Variables were screened for outliers using the SPSS Explore function (selecting outliers and descriptive statistics), which displays stem-and-leaf and boxplot distributions, amongst other data, to identify extreme outlier values. Outliers were replaced with a score that was one unit larger or smaller than the next most extreme score on the distribution (Tabachnick & Fidell, 2007). All variables were also screened for skewness and kurtosis levels. The three post-tour variables measuring frustration were severely skewed. Given that there was not a good theoretical rationale to transform these variables' scores, they were excluded from analyses that assumed a normal distribution. Of the 40 individual outlier scores that were cleaned across all variables, 15 were for frustration variables which were excluded from analyses. As such, 25 outlier scores (from 12 different variables) were both cleaned and used in our analyses. Other than the frustration variables, the one with the largest quantity of outliers was post-tour boredom directed toward learning about the Arts Building: Four outliers. All other variables had fewer or no outliers. See Appendix E and F for variables' descriptive statistics and correlations.

3.0. Results

3.1. RQ1: Did learners report feeling emotionally engaged during the tour? (a) What kinds of emotions did learners report experiencing toward different focal points? (b) Did learners' discrete emotions significantly differ based on the focal point that self-reported emotions were directed toward? (c) Did discrete emotions significantly differ between guide protocol conditions?

Descriptive statistics reported in Table 1 reveal relatively high mean levels of enjoyment and curiosity and low mean levels of frustration and boredom across focal points and conditions. In order to answer the second and third sub questions for RQ1, a repeated measures MANOVA was run where focal point (learning about the Arts Building, learning with the app, and guide questions) was entered as the three-level within subject variable, guide protocol condition as the two-level between subjects variable and enjoyment, curiosity, and boredom as the three dependent variables Frustration was not examined due to the extreme deviation from normality, though descriptive statistics and mean differences are reported in Table 1. Multivariate analyses revealed a significant main effect and large effect size of focal point on learners' emotions, Wilks' Lambda = .58, $F(6,38) = 4.62$, $p < .01$, $\eta^2_p = .42$, but no main effect of guide protocol or interact effect between guide protocol and focal point[7].

The follow-up repeated measures ANOVA measuring learners' tour-directed enjoyment revealed a significant main effect and medium effect size of focal point, $F(2,86) = 5.59$, $p < .01$, $\eta^2_p = .12$, no significant main effect of condition, and no

[7] Given that covariates only impact main and interaction effects (not within-subject), supplemental analyses using prior knowledge and gender (with the reduced sample size; see footnote 2 and 3) were not examined for non-significant main and interaction effects.

significant interaction effect between focal point and condition. Post hoc comparisons with a Bonferroni correction were used to determine if significant differences existed between different focal points (see Table 1). An examination of the descriptive statistics and comparisons revealed that learners self-reported significantly higher levels of enjoyment from learning about the history content (Arts Building) than interacting with the app, though both mean levels indicated that enjoyment was experienced.

The repeated measures ANOVA measuring learners' tour-directed curiosity revealed a significant main effect and medium effect size of focal point, $F(2,86) = 5.59$, $p < .01$, $\eta^2_p = .12$, no significant main effect of condition, and no significant interaction effect between focal point and condition. An examination of the descriptive statistics and pairwise comparisons revealed the same pattern as the results for enjoyment. No significant results were found in the repeated measure ANOVAs examining boredom.

3.2. (RQ2) Did learners in the EPF condition have higher (a) objective (post-tour history knowledge test score) and (b) subjective (perceived success) knowledge outcomes than those in the CPF condition?

In order to answer RQ2(a) a one-way ANOVA was conducted and revealed that learners in the EPF condition scored significantly higher on the post-tour test than those in CPF condition, $F(1,43) = 4.72$, $p < .05$, $\eta^2_p = .10$[8,9]. In order to answer RQ2(b) a one-way ANOVA was conducted and failed to reveal a significant main effect of condition on learners' perceived success, $F(1,43) = .407$, $p > .05$, $\eta^2_p = .02$, but the difference between

[8] Running an ANCOVA with a reduced sample (see footnote 1) and prior knowledge as a covariate failed to yield a statistically significant difference between conditions, but yielded a medium (though reduced) effect size, $F(1, 36) = 4.72$, $p > .05$, $\eta^2_p = .08$. The loss of power from reducing our sample by 13% to enter the covariate likely explains the loss of significance.

[9] Supplemental analyses with both prior history knowledge and gender entered as covariates yielded a marginally significant effect and medium effect size: $F(1,26) = 3.59$, $p = .07$, $\eta^2_p = .12$.

the two conditions did have a small effect size and matched the direction of the results from RQ2a (see Table 2)[10].

3.3. (RQ3) How did learners appraise task value? (a) Did task value differ between conditions? (b) Did learners report significantly different levels of task value before and post the guided tour?

In order to answer RQ3 and its two sub questions we ran a repeated measure ANOVA with condition as the independent variable and time (pre and post-tour) as the two-level within-subject variable for task-value. The repeated measures ANOVA measuring learners' tour-directed task value revealed a significant main effect of time, $F(1,42) = 5.93$, $p < .05$, $\eta^2_p = .12$. No significant main effect of condition was found, but a moderate effect size was observed, $F(1,42) = 2.35$, $p > .05$, $\eta^2_p = .05$. No significant interaction effect between time and condition was observed either. An examination of the descriptive statistics (see Table 3) revealed that learners reported significantly higher levels of task value after the guided tour when compared to their pre-guided-tour responses, although mean task value was high in both cases.

3.4. (RQ4) Did learners report higher levels of positive emotions and lower levels of negative emotions from learning about the Arts Building after the tour compared to the emotions they reported experiencing during typical (a) formal and/or (b) informal learning contexts?

To answer RQ4 a repeated measures MANOVA was run where learning context (Formal Contexts, Informal Contexts, and about the Arts Building) was entered as the three-level within subject variable, guide protocol condition as the two-level between

[10] A follow-up supplemental analysis examining the interaction between gender and condition on perceived success failed to yield a significant effect when prior knowledge was controlled for.

subjects variable and enjoyment and boredom as the two dependent variables. A paired t-test was conducted to examine differences between the level of frustration experienced in formal vs. informal contexts because of extreme deviation from normality for frustration experienced from learning about the Arts Building. Multivariate analyses revealed a significant main effect and large effect size of learning context on learners' emotions, Wilks' Lambda = .53, $F(2,39) = 8.72$, $p < .01$, $\eta^2_p = .47$, but no main effect of guide protocol or interaction effect between guide protocol and focal point[11].

The follow-up repeated measures ANOVA measuring learners' enjoyment of learning revealed a significant main effect of learning context, $F(2,86) = 17.45$, $p < .01$, $\eta^2_p = .29$. Post hoc comparisons with a Bonferroni correction were used to determine if significant differences existed between different contexts (see Table 4). An examination of the descriptive statistics and pairwise comparisons revealed that learners self-reported significantly higher levels of enjoyment from learning about the Arts Building in the guided tour than learning about history in formal settings. Learners also reported enjoying learning in informal settings significantly more than learning in formal settings, though not as much as learning about the Arts Building in the guided tour.

The repeated measures ANOVA measuring learners' boredom of learning revealed a significant main effect of learning context, $F(1.68,86) = 6.24$, $p < .01$, $\eta^2_p = .13$. Post hoc comparisons with a Bonferroni correction were used to determine if significant differences existed between different contexts (see Table 4). An examination of the descriptive statistics and pairwise comparisons revealed that learners self-reported significantly higher levels of boredom from learning in formal contexts compared to (1)

[11] Given that covariates only impact main and interaction effects (not within-subject), supplemental analyses using prior knowledge and gender (with the reduced sample size; see footnote 2 and 3) were not examined for non-significant main and interaction effects.

informal contexts, and learning about the (2) Arts Building in the guided tour. Learners also had significantly lower levels of boredom from learning about the Arts Building in the guided tour than learning about history in informal contexts.

A paired samples t-test revealed a significant difference between the level of frustration experienced in formal vs. informal contexts $t(43) = 2.81$, $p < .01$, Cohen's $d = 0.42$; the former being higher.

4.0 Discussion

This study had three aims: First and primarily, to investigate whether the EPF protocol supported better (a) emotional engagement, (b) knowledge outcomes, and (c) value of history learning than the CPF condition while learners used a mobile AR app to learn about history. The second aim of the study was to compare learners' appraisals of task value (learning about the history of McGill University) before and after the guided tour. The third aim of the study was to compare how learners felt about learning history during the guided tour with the mobile AR app compared to both typical formal and informal history learning contexts.

The main results of the study were the following: (1) Learners reported significantly higher levels of enjoyment and curiosity from learning about the Arts Building than from using the app itself, though mean levels indicated that enjoyment and curiosity were elicited in relatively high levels (in contrast to negative emotions) across the three focal points. (2) Knowledge measured using a post-tour test were significantly higher in the EPF than the CPF condition, though learners in both conditions scored highly. A small effect size was observed in the same direction for learners' perceived success. (3) Learners reported significantly higher levels of task value after the guided tour compared to their pre-guided-tour responses. (4) Learning context had a significant

effect on learners' emotions. Learners reported the highest mean level of enjoyment from learning about the Arts Building in the guided tour followed by learning about history in informal, then formal settings. The inverse pattern was true for boredom.

Results supported the hypothesis that learners would report high levels of enjoyment and curiosity and low levels of frustration and boredom from the guided tour. However, the hypotheses that there would be significant differences in the levels of emotions between conditions were not supported. Similar to our prior research (Harley et al., 2016a), learners' responses created a ceiling effect where positive emotions were so high and negative emotions so low that there was little variance remaining to observe differences between conditions. Despite the limited variation between responses, the hypothesis predicting significant differences between different focal points was confirmed. Specifically, learners reported significantly higher levels of enjoyment and curiosity when asked to report how they felt learning about the Arts Building versus interacting with the app. While learners tended to feel positively toward all aspects of the tour (e.g., all focal points), these results revealed that the learning content itself (learning about the Arts Building) contributed the most to their enjoyment and curiosity. This result makes sense, given findings that learners reported high task value toward learning about the history content and the relationship between high task value and positive emotions outlined in the CVT. Taken together, the results from RQ1 suggest that engaging in a tour designed to foster deeper and broader historical reasoning may not stimulate significantly higher positive emotions than the simpler CPF guide protocol, but learners none-the-less, reported that the guide's questions (and other focal points) elicited relatively high levels of enjoyment and curiosity across conditions. Additionally, the fact that (counter to our hypotheses) the historical learning content outshone technology as a source of positive

emotions is an important reminder that technology is a tool rather than an all-purpose solution for fostering emotional engagement. This is especially interesting in light of the paucity of research conducted examining emotions with mobile AR apps. It is possible, however, that the interaction with the historical content using the app may have highlighted its interest even though these questions aimed to separate focal points. Examining within-task discrepancies in emotions provides valuable information about what aspect(s) of a technology-rich learning environment may benefit from being adapted to support positive and beneficial emotional responses (Harley, Lajoie, Frasson, & Hall, 2017), especially in the absence of a control condition.

Our results supported the majority of our hypotheses pertaining to RQ2. Specifically, learners in the EPF condition scored significantly higher on the post-tour test than those in the CPF condition and learners in both conditions had high post-tour scores. These results provide preliminary evidence that both conditions were effective at teaching students about history with the mobile AR app. Moreover, these findings provide preliminary evidence that the guide's emphasis on a broader and deeper array of historical reasoning prompts in the EPF condition meaningfully enhanced learning over and above those prompts and feedback provided in the core CPF protocol (comparison-focused prompts shared by both).

Learners' perceived success was relatively high across conditions, as hypothesized, but did not vary significantly between conditions counter to hypotheses and expectations that perceived success should closely mirror objective learning outcome scores. It is possible that learners were hesitant to form stronger perceptions of success without having completed a formal test of their knowledge, which was administered immediately after the perception of success question.

Our set of hypotheses related to task value were confirmed. Specifically, learners reported high levels of task value both pre and post the guided tour, but significantly higher levels of task value post-tour. As we hypothesized, learners' motivation to participate may have been driven by an interest in learning about their university's history and task value may have increased post-tour as a result of learners believing (rather than speculating) that the material covered was interesting and valuable. These results converge with the high levels of positive emotions and low levels of negative emotions (see Appendix C).

The hypothesis that learners would report higher levels of positive emotions in informal contexts than formal was confirmed. This was not surprising because informal contexts are typically more self-selected (suggesting higher value) and involve more autonomy (suggesting higher control) than formal one's where choice is limited and appraisals of value are extrinsic (instrumental; task valued for grade) rather than intrinsic (the task itself is valued). The examples learners provided to situate their emotional rating of formal vs. information contexts supported these assumptions. The correlations (Appendix E) between emotions in informal contexts and emotions directed toward learning about the Arts Building also supported our hypotheses. These findings could suggest that learners may be motivated to use the mobile AR app to learn about history in informal, unstructured settings.

4.1. Limitations and Future Directions

One of the limitations of this study was the reliance on self-report data for measuring emotions. The use of single-item self-report measures to capture learners' emotions may also be seen as a limitation when compared to questionnaires with more items per scale that have been well established in the literature (e.g., AEQ; Pekrun et al.,

2002). There are, however, also advantages of self-reported measures and single-item ones, in particular. Single-item self-report questionnaires limit item fatigue, are less intrusive than questionnaires, and can also be more flexible. Asking learners to indicate how they feel about different focal points in a technology-rich learning environment (Conati & Maclaren, 2009; Harley et al., 2016a; Harley, Carter, Papaionnou, Bouchet, Azevedo, Landis, & Karabachian, 2016b), as this study did, is an example, and one that highlights an advantage of self-report measures of emotion over others: one cannot easily discern what elicited an emotion from behavioral or physiological data in open-ended learning environments (Harley, 2015). Studies have also found that the external validity of such measures is supported by behavioral measures of emotion (e.g., facial expressions; Harley et al., 2015), learning strategies and appraisals of task value (Pekrun et al., 2016) and volunteered expressions of emotion from session transcriptions (Harley et al., 2016a). Similar to Pekrun and colleagues (2016), the self-report measures of emotion used in this study are supported by correlations with task value in the expected directions. None-the-less, the addition of behavioral (e.g., facial expressions) and physiological measures of emotions (e.g., electrodermal activity) is an important future direction and one that is currently being pursued by the authors in a related app-based study.

Without a control condition, it is not possible to fully isolate the effect that the learning content, app, or guide protocol had on learners' emotions, appraisals, and knowledge outcomes. Therefore, while results from this exploratory study highlight the value of using self-report measures to capture more granular information about emotional responses directed toward specific focal points (e.g., Harley et al., 2016a,b), they should be seen as preliminary. While we had to prioritize committing our sample to investigating

the two different guide protocols, future research should seek to experimentally control for potential effects of using technology to better understand the relative role of external instructional scaffolds and learning content.

While the comparison in RQ4 between state emotions evoked during the study and emotional tendencies toward learning about history in formal and informal learning situations (trait emotions) was, as mentioned, not a pure one we believe it may add insight into the relationships between these two types of emotions and directions for future inquiry. Given the emerging empirical evidence investigating the boundaries and differences between trait and state emotions, results should be interpreted with some caution.

Another limitation of this study was the limited variance of some variables, particularly, the post-tour history knowledge test. These results were not surprising, however, given prior results from a similar study (Harley et al., 2016a,b) as well as other research indicating that learners at McGill University tend to score highly on learning outcome measures across experimental conditions and, in some observed cases, relative to other universities in cross-university studies (Harley, Bouchet, Papaionnou, Carter, Azevedo, & Landis, 2014). These repeated findings highlight the challenges of designing task-appropriate assessments for students at high-ranking international universities. This challenge is increased when the post-tour test is based on a limited amount of content and, therefore, a small number of items. The latter represents a limitation of the assessment of history knowledge in the study. Moreover, each item included in the post-tour test of this study was designed to assess different information and we therefore should not expect high internal reliability on the test (Rowe et al., 2017)—as other mobile AR app (and other types of technology-rich learning environment) researchers might find

should they examine the internal reliability of their tests (e.g., Wilson et al., 2018; Chang et al., 2015; Ibáñez et al., 2015; Li et al., 2013; Taub et al., 2018; Yoon et al., 2017). While others have reported low internal reliability (e.g., Chiang et al., 2014; Rowe et al., 2017), we do recognize that this limits our ability to make arguments regarding the reliability of the measure. Therefore, a direction for future research would be to expand on this first, short iteration of the knowledge test and include more items designed to test common concepts. In order to do so, future research should take learners on a larger tour that covers more locations, which could extend the database of questions and provide opportunities to observe more variance in scores as well as develop subscales where common types of knowledge can be grouped together. A larger repository of questions would also enable us to develop two equivalent tests that could be counter-balanced to provide both a formal, closed-ended pre and post-measure of learning. In the present study, learners' responses to two open-ended questions suggested that they knew very little about the historical content they interacted with. While the difference between responding to an open-ended vs. closed-ended, multiple-choice question is an imperfect measure of prior knowledge, it was further supported by many learners' admissions of "not knowing much".

As in many multimodal studies, and particularly those that take place outside of a laboratory, this study suffered data loss. Specifically, the loss of six participants' audio records (used for prior knowledge coding) and 18 participants' demographic data limited our ability to investigate the effect of prior knowledge and gender as covariates. Our supplemental analyses, which examined the effect of these covariates, when appropriate, suggests, however, that neither significantly impacted the effects we observed.

As this was an exploratory study, many other future directions remain. A prominent area for future inquiry is the examination of trade-offs between informal and formal learning environments which afford different levels of learner control. For example, the present study is an example of an informal and open-ended learning environment where learners had more control over how they navigated different locations and interacted with the learning material (i.e., in different orders and for different amounts of time.) In more formal and close-ended learning environments students typically have less control and their attention is more intentionally directed to aspects of the learning material for a specific amount of time. Future research should compare the affordances and limitations of these types of learning environments for student outcomes to allow learners and instructors to leverage the advantages under the appropriate circumstances.

Future research should also examine the relationship between historical reasoning and the emotional and appraisal variables this article examined; such analyses were beyond the scope of the current article. Finally, future research should also explore opportunities to integrate the guide protocol into an app so that a human guide is not needed and the app can be used to full effect in informal, learner-selected situations. Indeed, the first step to designing (and revising) adaptive and intelligent technologies such as these are often preliminary pedagogical validation studies on account of the expense and difficulties involved in automatizing a novel instructional framework (Azevedo, Greene, & Moos, 2007; Author, 2009) such as this one.

While this study does not offer context-specific guidelines for using the guide protocols, the app and guide protocol could (for example) be implemented in a classroom with a smart board replacing the physical tour location and students using an app to make

comparisons while the teacher supplied historical responding prompts. A previous study examined such a learning environment with the same app as this study and the same university student population (but not sample; Harley et al., 2016a). Finally, the authors are currently preparing to take a similar history app (that builds off of insights from this study) that addressing significant curricular gaps in history and social studies into schools; recommendations forthcoming.

4.3. Conclusions

In conclusion, this exploratory study builds upon prior work by Harley and colleagues (2016a) to provide preliminary support for the use of an extended (EPF) protocol that aims to foster a deeper and broader array of historical reasoning processes (Harley, Haldane, McLaughlin, Poitras, Lajoie…et al., 2018), primarily by posing questions to stimulate argumentation. The study replicated findings that learners enjoyed and experienced curiosity during the technology-mediated, guide-supported interaction while elaborating upon our understanding of why this might be: participants valued learning about the history of their university, something that only increased post-tour, and generally enjoyed learning about history (especially) in informal settings, and with the AR app. While learners did not report experiencing significantly greater levels of desirable emotions or task value in the EPF condition than the CPF condition, mean levels were high across guide protocol conditions. Preliminary results also suggest that the new protocol succeeded in fostering higher levels of knowledge; an important reminder that outcomes from instructional technologies may be enhanced by targeting pedagogical in addition to technological elements of the learning environment (Lajoie & Azevedo, 2006). Perceived success findings also provide preliminary support for this interpretation. Taken together, while preliminary, these findings paint a coherent and

optimistic picture regarding the use of mobile AR apps for teaching history while also providing guidelines for how history instruction can be paired with this emerging technology.

Acknowledgements: This research was supported by funding from the Social Sciences and Humanities Research Council (SSHRC) of Canada. The authors wish to thank Laura Pipe for assistance with data collection and Dr. Eric Poitras for contributions made to a previous study this one expanded on. The authors also wish to thank Philippe Latour for taking the picture used in Figure 2, Tara Tressel and Meredith Derian-Toth for posing as tour guide and learner, and Daniel Beaudin for graphics.

References

Azevedo, R., Greene, J. A., & Moos, D. C. (2007). The effect of a human agent's external regulation upon college students' hypermedia learning. *Metacognition and learning*, *2*(2-3), 67-87.

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

Bieg, M. (2013). *Trait and State Academic Emotions: Two Sides of the Same Coin?* (Doctoral dissertation). Retrieved from http://nbn-resolving.de/urn:nbn:de:bsz:352-253941

Chang, Y.L., Hou, H.T., Pan, C.-Y., Sung, Y.-T., Chang, K.E. (2015). Apply an augmented reality in a mobile guidance to increase sense of place for heritage places. *Educational Technology & Society*, *18*(2), 166–178.

Chang, H.Y., Hsu, Y.S., & Wu, H.K. (2016). A comparison study of augmented reality versus interactive simulation technology to support student learning of a socio-scientific issue. *Interactive Learning Environments*, *24*(6), 1148-1161.

Chen, C.H., Chou, Y.Y., & Huang, C.Y. (2016). An augmented-reality-based concept map to support mobile learning for science. *Asia-Pacific Education Researcher*, *25*(4), 567-578.

Chiang, T.H.C., Yang, S.J.H., & Hwang, G.J. (2014). An augmented reality-based mobile learning system to improve students' learning achievements and motivations in natural science inquiry activities. *Educational Technology & Society*, *17*(4), 352–365.

Conati, C., & Maclaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction*, *19*(3), 267-303.

Daniels, L. M., Haynes, T. L., Stupnisky, R. H., Perry, R. P., Newall, N. E., & Pekrun, R. (2008). Individual differences in achievement goals: A longitudinal study of cognitive, emotional, and achievement outcomes. *Contemporary Educational Psychology*, *33*(4), 584-608.

D'Mello, S. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology*, *105*(4), 1082.

Wilson, K. E., Martinez, M., Mills, C., D'Mello, S., Smilek, D., & Risko, E. F. (2018). Instructor presence effect: Liking does not always lead to learning. *Computers & Education*, 122, 205-220.

Eid, M., Schneider, C., & Schwenkmezger, P. (1999). Do you feel better or worse? The validity of perceived deviations of mood states from mood traits. *European Journal of Personality*, *13*, 283-306.

Elliot, A. J., & McGregor, H. A. (2001). A 2 x 2 achievement goal framework. *Journal of Personality and Social Psychology, 80,* 501–519.

Efstathiou, I., Kyza, E. A., & Georgiou, Y. (2018). An inquiry-based augmented reality mobile learning approach to fostering primary school students' historical reasoning in non-formal settings. *Interactive Learning Environments*, *26*(1), 22-41.

Frenzel, A. C., Pekrun, R., & Goetz, T. (2007). Girls and mathematics—A "hopeless" issue? A control-value approach to gender differences in emotions towards mathematics. *European Journal of Psychology of Education, 22*(4), 497.

Furió, D., Juan, M.-C., Seguí, I., & Vivó, R. (2015). Mobile learning vs. traditional classroom lessons: A comparative study. *Journal of Computer Assisted Learning, 31*, 189–201.

de Groot-Reuvekamp, M., Ros, A., & Van Boxtel, C. (2018). Improving Elementary School Students' Understanding of Historical Time: Effects of Teaching With "Timewise". *Theory & Research in Social Education*, *46*(1), 35-67.

Goetz, T., Bieg, M., Lüdtke, O., Pekrun, R., & Hall, N. C. (2013). Do girls really experience more anxiety in mathematics? *Psychological Science, 24*, 2079-2087.

Goetz, T., & Hall, N. C. (2013). Emotion and achievement in the classroom. In J. Hattie and E. M. Anderman (Eds.), *International guide to student achievement* (pp. 192-195). New York: Routledge.

Hall, N. C., Perry, R. P., Chipperfield, J. G., Clifton, R. A., & Haynes, T. L. (2006). Enhancing primary and secondary control in achievement settings through writing–based attributional retraining. *Journal of Social and Clinical Psychology*, *25*(4), 361-391.

Harley, J. M. (2015). Measuring emotions: A survey of cutting-edge methodologies used in computer-based learning environment research. In S. Tettegah & M. Gartmeier (Eds.). *Emotions, Technology, Design, and Learning* (pp. 89-114). London, UK: Academic Press, Elsevier.

Harley, J. M., Bouchet, F., Hussain, S., Azevedo, R., & Calvo, R. (2015). A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. *Computers in Human Behavior, 48,* 615-625.

Harley, J. M., Bouchet, F., Papaionnou, N., Carter, C., Azevedo, R., & Landis, R. (2014, April). Assessing learning with MetaTutor, a multi-agent hypermedia learning environment. In M.W. Chu, & J. M. Harley (Chairs), *Innovative practices for assessment in computer based learning environments.* Symposium conducted at the annual meeting of the American Educational Research Association, Philadelphia, PA.

Harley, J. M., Carter, C.K., Papaionnou, N., Bouchet, F., Azevedo, R., Landis, R. L., & Karabachian, L. (2016b). Examining the predictive relationship between personality and emotion traits and students' agent-directed emotions: Towards emotionally-adaptive agent-based learning environments. *User Modeling and User-Adapted Interaction, 26,* 177-219.

Harley, J.M., Haldane, C., McLaughlin, B., Poitras, E.G., Lajoie, S.P., Tressel., T., Pipe, L., Li, T., & Whittaker, A. (2018, April). *Beyond books, names, and dates: Using mobile augmented reality and prompts to foster historical reasoning.* Poster presented at the American Educational Research Association (AERA), New York, NY.

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

Harley, J.M., Poitras, E. G., Jarrell, A., Duffy, M. C., & Lajoie, S. P. (2016a). Comparing virtual and location-based augmented reality mobile learning: Emotions and

learning outcomes. *Educational Technology Research and Development, 64*(3), 359-388.

Huijgen, T., van Boxtel, C., van de Grift, W., & Holthuis, P. (2017). Toward historical perspective taking: Students' reasoning when contextualizing the actions of people in the past. *Theory and Research in Social Education, 45*(1), 110-144.

Huizenga, J., Admiraal, W., Akkerman, S., & Dam, G. T. (2009). Mobile game-based learning in secondary education: engagement, motivation and learning in a mobile city game. *Journal of Computer Assisted Learning, 25*(4), 332-344.

Hwang, G. J., Wu, P. H., Chen, C. C., & Tu, N. T. (2016). Effects of an augmented reality-based educational game on students' learning achievements and attitudes in real-world observations. *Interactive Learning Environments, 24*(8), 1895-1906.

Ibáñez, M. B., Di-Serio, Á., Villarán-Molina, D., & Delgado-Kloos, C. (2016). Support for augmented reality simulation systems: the effects of scaffolding on learning outcomes and behavior patterns. *IEEE Transactions on Learning Technologies, 9*(1), 46-56.

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.

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

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

Krathwohl, D. R. (2002). A revision of bloom's taxonomy: An overview. *Theory into Practice, 41*, 212-218.

Lajoie, S. P. & Azevedo (2006). Teaching and learning in technology-rich environments. In P. Alexander & P. Winne (Eds.), Handbook of Educational Psychology (2nd ed.) (pp. 803–821). Mahwah, NJ: Erlbaum.

Li, R., Zhang, B., Sundar, S. S., & Duh, H. B. L. (2013). Interacting with augmented reality: How does location-based AR enhance learning? In: Kotzé P., Marsden G., Lindgaard G., Wesson J., Winckler M. (eds) *Human-Computer Interaction – INTERACT 2013. Lecture Notes in Computer Science*, vol 8118. Springer, Berlin, Heidelberg.

Marcus, A. S., Stoddard, J. D., & Woodward, W. W. (2017). *Teaching history with museums: Strategies for K-12 social studies*. Taylor & Francis.

Mayer, R. E. (2002). Multimedia learning. *Psychology of Learning and Motivation*, *41*, 85-139.

Meinhardt, J., & Pekrun, R. (2003). Attentional resource allocation to emotional events: An ERP study. *Cognition and Emotion, 17,* 477–500.

Meyers, L., Gamst, G., & Guarino, A. (2013). *Applied multivariate research: Design and interpretation.* Thousand Oaks, CA: SAGE.

Pintrich, P. R. (1991). A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).

Pekrun, R. (in press). Achievement emotions. In A. J. Elliot, C. S. Dweck, & D. S. Yeager (Eds.), *Handbook of competence and motivation (2nd edition): Theory and application*. New York, NY: Guilford Press.

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., Elliot, A. J., & Maier, M. A. (2009). Achievement goals and achievement emotions: Testing a model of their joint relations with academic performance. *Journal of Educational Psychology, 101*(1), 115.

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. & Linnenbrink-Garcia, L. (2014). *International handbook of emotions in education.* New York, NY: Routledge.

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.

Pekrun, R., Vogl, E., Muis, K. R., & Sinatra, G. M. (2017). Measuring emotions during epistemic activities: the Epistemically-Related Emotion Scales. *Cognition and Emotion*, *31*(6), 1268-1276.

Plass, J. L. & Kaplan, U. (2016). Emotional design in digital media for learning. In S. Tettegah, & M. Gartmeier (Eds.), *Emotions, technology, design, and learning* (p. 131-161). New York, NY: Elsevier

Poitras, E. G., Harley, J. M., Compeasu, T., Kee, K., & Lajoie, S.P. (2016a). Augmented reality in informal learning settings: Leveraging technology for the love of history. In R. Zheng & G. Michael (Eds.). *Handbook of Research on Serious Games for Educational Applications* (pp. 272-293). IGI Global.

Rabina, D., & Cocciolo, A. (2012). Uncovering lost histories through GeoStoryteller: A digital GeoHumanities project. *Digital Humanities 2012*, Hamburg, Germany.

Rowe, E., Asbell-Clarke, J., Baker, R. S., Eagle, M., Hicks, A. G., Barnes, T. M., ... & Edwards, T. (2017). Assessing implicit science learning in digital games. *Computers in Human Behavior*, *76*, 617-630.

Ruthig, J. C., Haynes, T. L., Perry, R. P., & Chipperfield, J. G. (2007). Domain-specific optimistic bias: Implications for college students' achievement and well-being. *Social Psychology of Education, 10,* 115–137.

Scollon, C. N., Kim-Prieto, C., & Diener, E. (2003). Experience sampling: Promises and pitfalls, strengths and weaknesses. *Journal of Happiness Studies, 4,* 5-34.

Steptoe, A., Gibson, E. L., Hamer, M., & Wardle, J. (2007). Neuroendocrine and cardiovascular correlates of positive affect measured by ecological momentary assessment and by questionnaire. *Psychoneuroendocrinology, 32,* 56-64.

Stoel, G. L., van Drie, J. P., & van Boxtel, C. A. M. (2017). The effects of explicit teaching of strategies, second-order concepts, and epistemological underpinnings on students' ability to reason causally in history. *Journal of Educational Psychology*, *109*(3), 321-337.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson Education/Allyn and Bacon.

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

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.

Watson (2015) Emotions in the history museum. In H.R. Leahy, & S. Macdonald (Eds.). *The international handbook of museum studies.* Chichester, West Sussex [England]: John Wiley & Sons Ltd.

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.

van Boxtel, C., & van Drie, J. (2013). Historical reasoning in the classroom: What does it look like and how can we enhance it? *Teaching History*, 150, 44-52

van Boxtel, C., Grever, M., & Klein, S. (Eds.). (2016). *Sensitive Pasts: Questioning Heritage in Education* (Vol. 27). Berghahn Books.

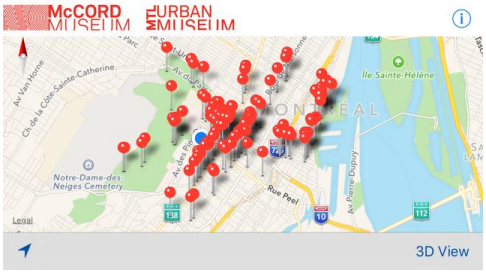
van Drie, J., & van Boxtel, C. (2008). Historical reasoning: Towards a framework for analyzing students' reasoning about the past. *Educational Psychology Review, 20,* 87-110.

van Drie, J., A. van Boxtel, C., & van der Linden, J. (2006). Historical reasoning in a computer-supported collaborative learning environment. In A.M. O'Donnell, C.E. Hmelo-Silver, & G. Erkens (Eds.). *Collaborative learning, reasoning, and technology* (pp.266-297. Mahwah NJ: Lawrence Erlbaum Associates.

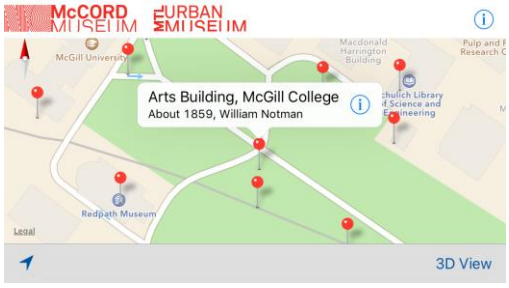
Yoon, S., Anderson, E., Lin, J., & Elinich, K. (2017). How augmented reality enables conceptual understanding of challenging science content. *Educational Technology & Society, 20*(1), 156–168.

Zhou, F., Duh, H. B. L., & Billinghurst, 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.

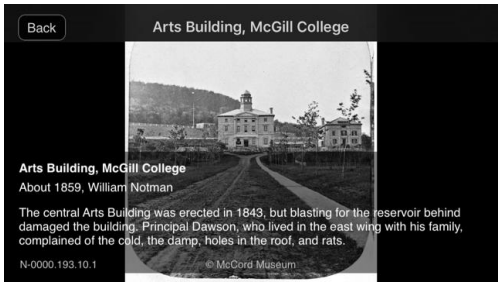
Caption 1



Caption 2



Caption 3



Caption 4

Figure 1. Interfaces of the MTL Urban Museum tours on the McCord Museum app used in the present study.

Figure 2. MTL Urban Museum tours on the McCord Museum app being used by learner (left) while interacting with tour guide (right) in front of contemporary version of the Arts Building. Image Photoshopped for illustrative purposes.

Table 1

Descriptive Statistics and Pairwise Comparisons for Focal Points by Condition

Emotion		Focal Points			Pairwise comparisons		
		Arts Building (1) *M(SD)*	App (2) *M(SD)*	Guide Questions (3) *M(SD)*	1&2 *MD(SE)[d]*	2&3 *MD(SE)*	1&3 *MD(SE)*
Enjoyment	All	4.33(.67)	3.96(.67)	4.04(.77)	.38**(.10).55	-.16(12)	.22(.12)
	EPF	4.32(.65)	3.87(.56)	3.86(.56)	-	-	-
	CPF	4.35(.71)	4.04(.77)	4.00(.80)	-	-	-
Boredom	All	1.58(.69)	1.62(.72)	1.44(.69)	-.04(.07)	.18(.11)	.13(.10)
	EPF	1.45(.60)	1.50(.67)	1.45(.74)	-	-	-
	CPF	1.69(.77)	1.74(.75)	1.74(.75)	-	-	-
Frustration	All	1.20(.41)	1.31(.60)	1.38(.68)	-.114(.09)	-.18(.11)	-.07(.10)
	EPF	1.18(.40)	1.41(.73)	1.5(.74)	-	-	-
	CPF	1.22(.42)	1.22(.42)	1.26(.62)	-	-	-
Curiosity	All	4.24(.74)	3.91(.93)	4.07(.75)	.34**(.10).49	.18(.09)	-.16(11)
	EPF	4.36(.73)	3.91(.97)	4.23(.75)	-	-	-
	CPF	4.13(.76)	3.91(.90)	3.91(.73)	-	-	-

Note. $p < .01$ = **; $p < .05$=*; *M* = mean; *SD* = standard deviation; *MD* = mean difference; *SE* = standard error; *d* = Cohen's *d*. Adjustment for multiple comparisons: Bonferroni; "-" = Not applicable. All emotions were measured using a five-point scale. EPF = extended prompt and feedback protocol. CPF = comparison-based prompt and feedback condition.

Table 2

History Learning Outcome Variables by Condition

	Condition				All	
Variable	CPF		EPF		*N*	
	M	*SD*	*M*	*SD*	*M*	*SD*
Post-tour Test Score	.79	.17	.88	.11	.83	.15
Perceived Success	7.39	1.25	7.68	1.25	7.53	1.52

Note. Post-tour score is a percentage. Perceived Success was on a 10-point scale. EPF = extended prompt and feedback protocol. CPF = comparison-based prompt and feedback condition.

Table 3

Descriptive statistics for task value by condition

		Pre-Guided Tour	Post-Guided Tour
	Condition	*MD(SD)*	*MD(SD)*
Task Value	EPF	3.96(1.11)	4.18(1.04)
	CPF	4.49(1.52)	4.79(1.04)
	Total	4.22(1.34)	4.49(1.25)

Note. EPF = extended prompt and feedback protocol. CPF = comparison-based prompt and feedback condition.

Table 4

Descriptive Statistics and Comparisons for Emotions Toward Different Focal Points

Emotion	Learning contexts			Pairwise comparisons		
	Formal (1) *M(SD)*	Informal (2) *M(SD)*	Arts Building (3) *M(SD)*	1&2 *MD(SE)d*	2&3 *MD(SE)d*	1&3 *MD(SE)d*
Enjoyment	3.57(.99)	4.14(.82)	4.32(.67)	.57**(.13).65	.18(11)	.750**.76
Boredom	2.04(1.06)	1.80(1.10)	1.60(.69)	.25*(.10).38	.21(.15)	.46**(.14).50
Frustration	1.91(.1.00)	1.57(.90)	1.20(.41)	.34(.12)*.42	.36(.15)	.71(.15)

Note. $p < .01$ = **; $p < .05$=*; *M* = mean; *SD* = standard deviation; *MD* = mean difference; *SE* = standard error; *d* = Cohen's *d*. Adjustment for multiple comparisons: Bonferroni; "-" = Not applicable. All emotions were measured using a five-point scale.

Appendix A

Significance testing for pre-tour variables between randomly assigned conditions

Variable	Mean (Standard Deviation)		Independent samples t-test
	EPF condition	CPF condition	
Task-value Global (Pre)	3.96(1.10)	4.49(1.52)	$t(42) = 1.31$, $p > .05$
Pre-tour knowledge	.11(.09)	.08(.09)	$t(37) = -.86$, $p > .05$
Enjoy Pre-tour (Formal)	3.50(.91)	3.63(1.09)	$t(42) = 0.45$, $p > .05$
Bored Pre-tour (Formal)	1.95(1.09)	2.14(1.04)	$t(42) = 0.57$, $p > .05$
Frustration Pre-tour (Formal)	1.81(1.05)	2.00(.87)	$t(42) = 0.62$, $p > .05$
Enjoy Pre-tour (Informal)	4.09(.81)	4.18(.85)	$t(42) = 0.36$, $p > .05$
Bored Pre-tour (Informal)	1.59(.85)	2.00(1.27)	$t(42) = 1.25$, $p > .05$
Frustration Pre-tour (Informal)	1.50(.80)	1.64(1.00)	$t(42) = 0.50$, $p > .05$

Note. The pre-tour measure of knowledge variable had the same possible range as the post-tour knowledge quiz 0 – 1.00. Please see Appendix F for descriptive statistics of other study variables.

Appendix B

Summary of guide prompts and feedback by category and type

Prompt and Feedback Category	Guide Protocol Condition	Prompt and Feedback Type	Prompt and Feedback Description
General	Both	Elaborate/content pump	"OK" or "right" + "can you tell me more?"
		Metacognitive pump	What are you basing that claim on? What prior knowledge are you drawing on when you say that? How confident are you about X?
		Prompt (for answer)	Any thoughts? Notice anything? Any come to mind?
		Agreement / positive feedback	Yes, that's right.
		Clarify	I'm not quite sure I understand what you mean. Could you try to explain it again?
Scheduled	EPF	Hint	See second column of Appendix B E.g., *Why might these have been added?*
	Both	Provide/ splice in Answer	See first column of Appendix B E.g., *McGill has grown*
	Both	Provide historical context	See third column of Table 2 E.g., The *faculty consisted of one Principle and four professors in 1821 when first founded (but before it formally opened). As of fall 2014 there were 1,674 full time professors*
Returning/Reset	Both	Open-ended introduction question	Comparing the historical information on the Arts Building and its surroundings presented in the app to that of the present-day, do you notice any differences or similarities? Please use the app to view additional, nearby locations—the red markers like the one you clicked on—to help you.
	Both	Procedural clarifications	Please focus your comparison on the Arts Building and its surrounding area that the historical image is of, as if you were using it as a window to look back in time. This tour won't be focusing on similarities and differences across the whole campus, just that area of interest.

Note. Protocol-specific guide prompts are labelled as "scheduled" while protocol-general are labelled as "general" and "returning/reset" as per the second column notation (EPF = extended prompt and feedback protocol only). This appendix makes reference to Appendix C which provides an alternative representation and example of scheduled prompts related to one of the historical differences in the tour.

Appendix C

Historical differences between past and present and corresponding prompts, feedback, and historical information

Historical differences	Historical reasoning hints (EPF only)	Context/Meaning of difference	Additional historical context and facts about difference
A1. No east (Dawson Hall) and west (Molson Hall) wing for the Arts Building.	H. Why might these have been added?	A: McGill has grown	HC1: Faculty: one Principle and four professors in 1821 when first founded (but before it formally opened). As of fall 2014 there were 1,674 full time professors HC2. On McGill's first day of classes in 1843 there were only twenty students. As of fall 2014 there were 39,500

Note. Alternative representation and example of scheduled prompts related to one of the historical differences in the tour. Includes: (column A) one of the historical differences they must identify using the historical multimedia from the mobile AR app and comparisons with the current day version of the location; (column B) the hint a learner will receive after identifying the corresponding historical difference if they are in the extended prompt and feedback (EPF) condition; (column C) the context and meaning behind the historical difference which is the answer the guide will provide them with if they are in the comparison-based prompt and feedback condition (CPF). Alternatively, if they are in the EPF protocol condition the guide will try to encourage them to reach the meaning of the difference themselves, potentially using the general prompts from Appendix B (which may also be used in helping them identify historical differences in the first place—answer dependent). After the meaning of the difference has been identified or disclosed the guide then provides corresponding additional historical context and facts about the difference that supplement the historical multimedia information. Please see 2.3. for more details.

Appendix D

Post-tour quiz questions and answers

Question	Options

1. How many students were enrolled at McGill on its first day of classes in history?	(a) 5 (b) 20* (c) 100 (d) 2000 (e) 3000
2. What is McGill University's most historical symbol / icon?	(a) The Roddick Gates (b) Molson Hall (c) Dawson Hall (d) The Arts Building's cupola (and flagpole)* (e) James McGill monument
3. When did McGill University first open its doors for classes?	(a) September 6, 1843* (b) September 1, 1902 (c) September 9, 1899 (d) September 4, 1820 (e) September 2, 1936
4. Who founded McGill University?	(a) Sir John William Dawson (b) Sir Wilfrid Laurier (c) James McGill * (d) Reverend George Jehosophat Mountain (e) Francis Desrivières
5. Where was the founder of McGill University born?	(a) Glasgow, Scotland* (b) Ottawa, Canada (c) Montréal, Québec (d) London, England (e) Paris, France
6. Which of the below may have hinted at the founder of McGill's future interest in founding a university? He/She:	(a) Chartered a law mandating government support for higher education (b) Led the defence of Montreal during the War of 1812. (c) Made his fortune as a successful merchant, trading fur, ammunition, and general goods. (d) Taught English as a second language (e) Drafted a petition calling for the improvement in education*
7. How did the founder found McGill University?	(a) Conquered the land the university is built upon. (b) Married a Francophone widow, Charlotte, and petitioned her father for the funds.

(c) Used his position as a member of the Legislative Assembly for the colony of Lower Canada to acquire a Crown endorsement.
(d) Bequeathed in trust £10,000 and his forty-six-acre Burnside Place to the cause.*
(e) Petitioned the Royal Institution for the Advancement of Learning for the funds and land.

Note. * = correct answer.

Appendix E

Correlations between study variables

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Post-tour knowledge test	1.00																						
Perceived Success	-.06	1.00																					
Enjoy Arts Building	-.07	.33*	1.00																				
Enjoy App	.03	.40**	.48**	1.00																			
Enjoy Guide	-.07	.43*	.33*	.37*	1.00																		
Bored Arts Building	.05	.00	-.52**	-.09	-.13	1.00																	
Bored App	-.10	-.08	-.27	-.14	-.08	.77**	1.00																
Bored App	-.10	-.08	-.35*	-.22	-.13	.77**	.94**	1.00															
Bored Guide	.19	-.01	-.373*	-.25	-.19	.50**	.37*	.44**	1.00														
Frustrated Arts Building	.30*	-.07	-.58**	-.13	-.15	.64**	.42**	.50**	.41*	1.00													
Frustrated App	.03	-.04	-.09	-.08	-.13	.27	.34**	.49**	.32*	.40**	1.00												

Frustrated Guide	.08	.17	-.23	-.31*	-.04	.20	.31*	.34*	.69**	.21	.43**	1.00											
Curiosity Arts Building	.14	.43**	.42**	.48**	.36*	-.15	-.16	-.21	-.13	-.09	.13	.04	1.00										
Curiosity App	.01	.44**	.49**	.69**	.48**	-.17	-.25	-.33*	-.33*	-.19	-.20	-.23	.69**	1.00									
Curiosity Guide	.01	.37*	.22	.50**	.51**	-.08	.03	.01	-.19	-.04	.16	-.05	.70**	.60**	1.00								
Task-value Global (Post)	-.22	.47**	.62**	.57**	.49**	-.04	.03	.11	.16	.21	.03	.16	.64**	.69**	.46**	1.00							
Task-value Global (Pre)	-.17	.53**	.67**	.50**	.47**	-.07	.01	.11	.09	.27	.09	.12	.54**	.62**	.32*	.86**	1.00						
Enjoy Formal (Pre)	.05	.35*	.35*	.13	.24	-.40**	-.40**	-.39**	-.15	-.24	-.15	-.23	.23	.16	.10	.26	.374**	1.00					
Bored Formal (Pre)	-.17	.06	-.28	.10	-.10	.51**	.40**	.39**	.10	.14	-.06	.04	.05	.07	.03	.10	-.03	-.58**	1.00				
Frustrated Formal (Pre)	-.04	-.06	-.21	-.01	-.15	.22	.21	.32*	.06	.11	-.03	.09	-.13	-.09	-.06	-.08	-.22	-.53**	.62**	1.00			
Enjoy Informal (Pre)	.17	.33*	.51**	.27	.32*	-.10	.02	-.03	.01	-.02	.19	-.05	.40**	.32*	.21	.40**	.50**	.55**	-.49**	-.51**	1.00		
Bored Informal (Pre)	-.23	-.10	-.19	.21	.01	.47**	.46**	.41**	-.06	.04	.00	-.08	.06	.10	.10	.15	.05	-.55**	.82**	.52**	-.41**	1.00	
Frustrated Informal (Pre)	-.24	-.11	-.31*	.03	-.08	.42**	.44**	.40**	-.05	.06	-.04	.01	.01	-.05	.08	.03	-.16	-.55**	.78**	.63**	-.52**	.81**	1.00

*. = Correlation is significant at the 0.05 level (2-tailed); ** = Correlation is significant at the 0.01 level (2-tailed)

Appendix F

Descriptive information of study variables

Variable	*N*	Possible Range	Actual Range	Min	Max	*M*	*SD*
Post-tour History Knowledge Test	45	0-.1.00	0.43	0.57	1.00	0.84	0.14
Perceived Success	45	1-10	5.00	5.00	10.00	7.53	1.52
Enjoy Arts Building	45	1-5	2.00	3.00	5.00	4.33	0.67
Enjoy App	45	1-5	2.00	3.00	5.00	3.96	0.67
Enjoy Guide	45	1-5	2.00	3.00	5.00	4.11	0.75
Bored Arts Building	45	1-5	2.00	1.00	3.00	1.58	0.69
Bored App	45	1-5	4.00	1.00	5.00	1.69	0.90
Bored App	45	1-5	2.00	1.00	3.00	1.62	0.72
Bored Guide	45	1-5	2.00	1.00	3.00	1.44	0.69
Frustrated Arts Building	45	1-5	1.00	1.00	2.00	1.20	0.40
Frustrated App	45	1-5	2.00	1.00	3.00	1.31	0.60
Frustrated Guide	45	1-5	2.00	1.00	3.00	1.38	0.68
Curiosity Arts Building	45	1-5	2.00	3.00	5.00	4.24	0.74
Curiosity App	45	1-5	3.00	2.00	5.00	3.91	0.92
Curiosity Guide	45	1-5	2.00	3.00	5.00	4.07	0.75
Task-value Global (Post)	45	1-7	4.60	2.40	7.00	4.49	1.25
Task-value Global (Pre)	44	1-7	5.40	1.60	7.00	4.23	1.34
Enjoy Formal (Pre)	44	1-5	3.00	2.00	5.00	3.57	1.00
Bored Formal (Pre)	44	1-5	4.00	1.00	5.00	2.05	1.06
Frustration Formal (Pre)	44	1-5	3.00	1.00	4.00	1.91	0.96
Enjoy Informal (Pre)	44	1-5	3.00	2.00	5.00	4.14	0.82
Bored Informal (Pre)	44	1-5	4.00	1.00	5.00	1.80	1.09
Frustration Informal (Pre)	44	1-5	3.00	1.00	4.00	1.57	0.90