Facial expressions when learning with a Queer History App: Application of the Control Value Theory of Achievement Emotions

Authors and Biography

Byunghoon “Tony” Ahn, Jason M. Harley

Byunghoon “Tony” Ahn is a PhD student in the Experimental Surgery program at McGill University, where he studies surgical education. His interests include applied design, psychology for medical education. Dr. Jason M. Harley is an Assistant Professor at the Department of Surgery, McGill University; a Junior Scientist at the Research Institute of the McGill University Health Centre (RI-MUHC); an Associate Member of the Institute for Health Sciences Education, McGill University; and an Adjunct Professor at the Department of Educational Psychology, University of Alberta. Their interests include surgical education, emotion, technology, and LGBTQ+ issues in education.

Address for correspondence: Byunghoon “Tony” Ahn, Department of Surgery, Montreal General Hospital, 1650 Cedar Ave, L3-315, QC, H3G 1A4. *Email:* tony.ahn@mail.mcgill.ca

Abstract

Learning analytics (LA) incorporates analyzing cognitive, social, and emotional processes in learning scenarios to make informed decisions regarding instructional design and delivery. However, there is a gap in research investigating the relationship between direct behavioral traces of emotions and learning. We contributed to addressing this gap by examining learners’ facial expressions while they interacted with a multimedia mobile app to learn about queer history in a North American city. Specifically, we used automatic facial recognition software (FaceReader 7) to measure learners’ discrete emotions, an eye-tracker to identify emotions experienced while learners read specific content versus over the course of the entire learning session, and a counter-balanced multiple-choice quiz to assess learning. Results revealed that learners expressed more negative-activating emotions (i.e., anger, anxiety) and negative-deactivating emotions (i.e., sadness) than positive-activating emotions (i.e., happiness) and learners with an angry emotion profile had the highest learning gains. The importance of examining typically undesirable emotions in learning, such as anger, is discussed using the control-value theory of achievement emotions. Further, this study describes a multimodal methodology to integrate behavioral trace data into learning analytics research.

Keywords: Emotion, Learning Analytics, Multimodal, Multimedia, Educational Psychology, LGBTQ

Structured Practitioner Notes

What is already known about this topic

- Learning analytics is about collecting and analyzing data of learning scenarios to understand learners, teachers, and their context, for enhancing learning experiences.
- Multimodal analytics have increasingly gained traction, with researchers incorporating more advanced methodologies and tools for their studies.
- Emotions play a critical role in learning, impacting learners' cognitive, motivational, and regulatory processes.
- The Control Value Theory of Achievement Emotion predicts positive-activating emotions (e.g., enjoyment) should lead to better performance.

What this paper adds

- Application of Control Value Theory of Achievement Emotion on less studied subject domain (i.e., history) and matter (i.e., queer culture).
- Preliminary evidence that negative-activating emotions (e.g., anger) can facilitate learning in certain contexts, and that emotions may play different roles depending on the subject matter and domain.
- Insights into methods of aligning facial recognition data from facial expression analysis software with eye-tracking data dealing with dynamic content.

  - Insights into frequency and the types of emotions elicited in learning scenarios dealing with sensitive topics

Implications for practice and/or policy

- Educators should be aware of different types of emotions, and their roles in learning scenarios.
- Educators should critically evaluate whether emotions with positive valence always have a positive impact on learning, and vice versa.
- Learners may not always behaviorally express emotions through facial expressions, including when gazing at sections of learning material directly connected to assessment; therefore, it is helpful to supplement granular with larger timeframe analyses to examine emotional profiles.

Introduction

Learning analytics (LA) is commonly defined as "the measurement, collection, analytics and reporting of data about learners and their contexts, for the purposes of understanding and optimizing learning and the environments in which it occurs" (Long & Siemens, 2011, p.34; Viberg, Hatakka, Bälter, & Mavroudi, 2018). One core aim of LA is therefore to understand the factors that impact learning. Toward this end, LA researchers have embraced various technologies and techniques for collecting and analyzing data (Nistor & Hernández-García, 2018), including an approach of combining data from multiple sources: multimodal data analysis (Di Mitri, Schneider, Specht, & Drachsler, 2018). Di Mitri and colleagues report that the combination of steady development and interest of multimodality over the past two decades, and the recent technological advancements (e.g., wearable sensors, etc.) have proliferated multimodal research.

Despite the advancements of LA and multimodal analysis, there are still gaps in the literature when it comes to examining emotions in relation to learning outcomes, especially using alternative methods to self-reports (e.g., Di Mitri et al., 2018). We present a study that uses multimodal data to investigate emotions and learning outcomes. Specifically, we examined learners' emotions in real-time with automatic facial recognition software while they used a multimedia mobile app to learn about queer history. Eye-tracking was also used for one of the analyses to infer learners' attention to specific content related to the administered knowledge-test questionnaire. Our article contributes to the literature by utilizing a framework apt for studying academic emotions (Control Value Theory of Achievement Emotions; Pekrun, 2006), while applying and expanding on previous methodologies towards a subject domain/matter seldom studied: queer history. We further include appendices detailing the novel approach we developed for leveraging eye-tracking to provide us with detailed contextual information during learners key moments of experienced emotions.

Achievement Emotions and Learning

Pekrun's (2006) Control Value Theory (CVT) of Achievement Emotions was used to help us formulate research questions, and as a critical lens to interpret results. Achievement emotions refer to emotions related to achievement activities or outcomes (e.g., anxiety during tests; Pekrun, 2006). Two cognitive appraisal processes act as antecedents for such emotions: 1) the perceived level of control, and 2) the level of value attributed towards learning. A student delivering a well-prepared presentation to pass a course is likely to have high appraisal levels of control and value. In general, high control appraisals are associated with positive emotions, and vice versa (Pekrun, 2011). Similarly, high perceptions of value are associated with more intense emotions, and vice versa. Achievement emotions can be categorized according to a three-dimensional taxonomy: valence (negative or positive), activation (deactivating or activating), and object focus (retrospective outcome, concurrent activity, or prospective outcome). For example, delivering a well-prepared presentation may elicit enjoyment—an emotion that would be categorized as having positive value, high activation, and having an object focus on the concurrent activity.

Achievement emotions can impact learners' cognitive, motivational, and regulatory processes, and are associated with learning outcomes (Pekrun & Perry, 2014). The CVT predicts that positive-activating emotions (e.g., enjoyment) should lead to better learning outcomes, while the inverse is true for negative-deactivating emotions (e.g., hopelessness). Positive-deactivating emotions (e.g., relaxation) and negative-activating emotions (e.g., anger) yield mixed results, but emotions such as anger should lean toward lower performance on average (Pekrun, Lichtenfeld, Marsh, Murayama, & Goetz, 2017). These predictions are based on mediating cognitive processes, and how emotions can impact these processes (Pekrun 2006; Pekrun & Perry, 2014). For example, feeling hopeless toward studying for a test may lead to superficial studying strategies (e.g., simple repetitive rehearsal) and off-task behavior (e.g., worrying about failure). On the other hand, feeling enjoyment during learning may lead to better regulation strategies (e.g., formulating achievement goals) and facilitate cognitive processes (e.g., activation of working memory resources).

Measuring Emotions: Facial Recognition Software

The expressive component of emotions includes facial expressions formed by activities of muscle groups called action units (Ekman & Friesen, 1978). Recent methodological development has led to trained artificial neural networks that can match varying sets of action unit activities to different discrete emotions often based on Ekman’s (1992) program of research and theory of basic. The theory of basic emotions proposes that emotional facial expressions are universal due to emotions being a product of evolution. Experts and software rely on this proposition and the action unit activities to code emotions with a high level of reliability. The theory of basic emotions lists six universal emotions: anger, fear, disgust, sadness, happiness, and surprise. These can be grouped into categories proposed by the CVT: anger, fear, and disgust are negative-activating emotions; sadness is a negative-deactivating emotion; happiness is a positive-activating emotion. Surprise can be categorized as a non-valenced (neutral) activating emotion, due to it being prone to learning towards either direction of the valence spectrum (Harley, Bouchet, & Azevdeo, 2013; Harley, Bouchet, Hussain, Azevedo, & Calvo, 2015).

While studies often employ expert coders to detect facial expressions (Di Mitri et al., 2018), this methodology has disadvantages such as potential for interrater reliability issues, and the amount of time, financial, and human resources it demands. Facial recognition software can address these issues, while yielding comparably accurate results. In recent years, facial recognition software has achieved higher reliability with grounded truth measures (e.g., self-report) relative to other methods such as physiological measures (Harley, 2016). Moreover, previous studies have utilized facial recognition technology when examining emotions in achievement situations similar to this study (e.g., Harley et al., 2013, 2015).

Related Literature

Emotions have been examined in various learning scenarios, including ones involving medical education (Artino, Holmboe, & Durning, 2012a, 2012b) and multimedia learning (Stark, Malkmus, Stark, Brümlem, & Park, 2018; Um, Plass, Hayward, & Homer, 2012). However, many such studies lack incorporating a theoretical framework built to deal with academic emotions. Popular theories in education research such as cognitive load theory (CLT; Sweller, 1988) and cognitive theory of multimedia learning (CTML; Mayer, 2005), for example, lack consideration towards emotions in learning (Um et al., 2012). Applications of these theories tend to over-simplify emotions as contributors to extraneous load that ultimately hinder learning by taking up limited cognitive resources (Harp & Mayer, 1998; Um et al., 2012). Further, the extended version of CTML, the cognitive-affective theory of media learning (CATML; Moreno & Mayer, 2007), while acknowledging affective factors, does not formulate propositions about differentiating discrete emotions (e.g., general positive emotions vs. learning-related positive-activating emotions; Stark et al., 2018). The lack of differentiation has led to the tendency of lumping emotions into a simplified dichotomous category of positive or negative. For example, previous studies may group emotions such as “upset”, “hostile”, and “scared” into a single category of “negative affect”, when using measures such as the Positive and Negative Affect

Schedule (PANAS-SF; Watson, Clark, and Tellegen's, 1988). One of the symptoms of this limitation is varying inconsistent findings in the impact of emotions in learning, and the inability to explain the inconsistencies (Stark et al., 2018).

This study addresses the issue of examining emotions within the framework that directly deals with academic emotions (i.e., the CVT), by drawing on findings and methodologies from previous empirical research that have examined emotions using CVT in various technology rich settings. Our study further contributes to the literature by examining a topic rarely looked at within educational research: queer history. Such topics warrant examination as education can contribute to preventing homophobia and transphobia (Harley, Liu, et al., 2019), all the while confirming generalizability of previous studies to a relatively novel learning scenario.

There are several related literatures we drew our methodologies from. First, Harley and colleague's (2015) study examined students learning about the human circulatory system in an intelligent hypermedia learning environment. The study focused on evaluating the synchronous use of three data channels when measuring emotions: facial recognition technology, self-reports, and electrodermal activity measurements. While the study revealed a high agreement rate (75.6%) between self-reports and facial recognition, it also noted the general dominance of neutral emotions in participants. The study further contributed to strengthening the validity of automatic facial recognition software when measuring emotions during learning activities.

Taub et al (in press) have used MetaTutor to investigate the relationships between emotions from facial recognition software, cognitive and regulatory processes, and learning outcomes. The authors highlighted findings that surprise and feeling of knowing (metacognitive judgment) were negatively correlated and that frustration was positively correlated with accurate note taking (cognitive learning strategy). The latter result was contrary to their predictions, but they suggested that learners who experienced frustration, were perhaps better engaged and motivated to take the most informative and accurate notes as the source of frustration was from not fully understanding the learning material. The authors noted that the helpful potential of negative emotions such as anger and frustration should be further explored.

Another relevant study comes from Jarrell, Harley, and Lajoie (2016), who examined achievement emotions in BioWorld, a software to support medical diagnostic reasoning. They found that 50% (13) of their learners were categorized as low emotion learners, while 26.9% (7) and 19.2% (5) were categorized as positive emotion learners and negative emotions learners, respectively. The low emotions learners seemed to lack the experience of high intensity emotions, and hence expressed less emotions relative to other groups. Despite this, low emotion learners still reported moderate levels of enjoyment towards learning. Results also revealed that while participants experiencing positive emotions reported the highest appraisals of task value and control, they had lower learning performance compared to learners with low emotions, and had comparable results to learners with negative emotional profiles. The authors explained this result by pointing out the possibility of emotions taking up cognitive resources that should have been devoted to learning.Harley et al. (2019) used the same app used in this study, although we asked different research questions and used a different sample. The study results included learners reporting high mean values of enjoyment, and low mean values of boredom. Furthermore, Harley, Poitras, Jarrell, Duffy, and Lajoie (2016) in a separate, but similar study with a different mobile app looked at emotions and learning outcomes in a guided history tour. They found that 46% of the learners reported enjoyment, and only 8% of the learners expressed a negative emotion towards learning with the app. An extension of this study with a new and larger sample (Harley, Poitras, Jarrell, Duffy, & Lajoie, in press) examined the effect of two different guide-facilitated historical reasoning protocols on emotional engagement, knowledge outcomes, and value of history learning. Results indicated both protocols supported positive emotions such as enjoyment and found low mean levels of negative emotions. Both protocols led to high knowledge outcomes, although the extended prompt and feedback protocol led to better knowledge outcomes compared to the latter. Finally, Poitras, Harley, and Liu (2019) examined DiscoverUofU, a location-based AR mobile app that was used to give a historical guide of a university campus. The results indicated that students could be clustered into groups with positive and negative emotional profiles. Learners who experienced enjoyment in learning with the app were more successful at identifying distractor information during measurements of topic understanding.

D'Mello (2013) reported a meta-analysis that focuses on identifying sets of discrete emotions that are found across various learning scenarios furnished with learning technologies. The study analyzed 24 studies, featuring diverse education level (middle school to adult education), ethnicity/country (e.g., UK, Philippines, etc.), sample size (*M* = 73, *SD* = 66), learning settings (e.g., laboratory settings), technologies (e.g., simulations), and topics (e.g., microbiology). The meta-analysis revealed that engagement/flow, boredom, confusion, curiosity, happiness, and frustration seemed to be commonly found in learning scenarios with technology. Further, affects were found to be influenced by the activity, location, and source of the measurements: Engagement/flow was associated with advanced learning technologies (e.g., serious games), authentic learning contexts, and observer reports; frustration and boredom was associated with laboratory studies, more basic technologies (e.g., simple web interface), and self-reports.

The above studies, while different, highlight common grounds that helped us formulate hypotheses for our research questions. Specifically, the studies have shown that while negative emotions may show potential in facilitating learning, most studies indicate that positive emotions are the most beneficial for supporting learning. Further, learning situations involving mobile multimedia apps for learning history seem to elicit enjoyment of learning from the learners in general, other emotions may be captured depending on authenticity of the learning scenario and the level of sophistication of the learning technology used. Finally, the studies make a case for facial recognition software being a valid method for measuring emotions in real time.

Research Questions and Objectives

To investigate the impact of emotions in learning outcomes we carried out two sets of analyses guided by three research questions: (RQ1a) how often did learners express emotions while they read content directly related to the answers on the knowledge check questionnaire, and (b) which discrete emotions were expressed? The second analysis aimed to answer the following: (RQ2a) How often did learners express emotions during the entire learning sessions, and (b) which discrete emotions were expressed? (RQ3) Were there statistically significant group differences in proportional learning gains when learners were grouped by their dominant emotion (e.g., anger vs. happiness) during the entire learning session?

Hypotheses. The CVT and previous research that examined emotions with mobile apps played a particularly important role in forming hypotheses for the above research questions. Harley et al. (2015) provided evidence that neutral emotions tend to be a dominant emotion for learners when emotions are analyzed using automatic facial recognition software and self-report measures. Jarrell et al. (2016) have corroborated this finding that many learners tend to express low levels of emotion. We have therefore hypothesized for RQ1a and RQ2a that we would expect learners to express neutral emotions most of the time during learning.

Previous findings with the Queer History App used in this study as well as other similarly designed apps for learning history (Harley et al., 2016; 2019, in press; Poitras et al., 2019) indicate that high mean levels of emotions, such as enjoyment, and low mean levels of negative emotions, such as frustration and boredom, are experienced with such apps. Based on these reports, we have hypothesized that learners would tend to express happiness (positive-activating), and tend to express little sadness (negative-deactivating), or anger (negative-activating) for RQ1b and RQ2b.

Lastly, the CVT literature has consistently shown that, on average, positive-activating emotions such as enjoyment (likely expressed as happy in terms of facial expression) will have a positive impact on learning outcomes (Pekrun & Perry 2014; Loderer, Pekrun, & Lester et al., in press). Negative-deactivating emotions such as boredom (likely expressed as sad in terms of facial expression; Harley et al., 2015), on the other hand, will have a negative impact. While negative activating emotions such as anger are proposed to have more complex results, on average, it is reasonable to expect negative outcomes, although to a lesser extent relative to negative-deactivating emotions. Based on these guidelines we have hypothesized that for RQ3, the learners with positive-activating dominant emotions (i.e., happy) would have higher proportional learning gain scores compared to learners with other emotions. We also hypothesized that learners with negative-activating dominant emotions (i.e., anger, scared) would perform better than learners with negative-deactivating dominant emotions (i.e., sad), but would perform worse than learners with positive-activating dominant emotions.

Methods

Participants

The current study's data comes from a larger study with 114 pre-service teachers (33.3% male sex; 77 Caucasian) from a large North American university. They were 18 to 41 years old (*M* = 23.1 years, *SD* = 4.86) with a self-reported GPA range of 2.00 to 4.00 (out of 4.00; *M* = 3.11, *SD* = 0.39, four missing values). Participants were recruited from an undergraduate education course where they chose between experiments or an alterative activity to earn course credits. They were randomly assigned to two conditions, where one was given a mobile multimedia app on queer history, and the other was given a game to play (Plants vs. Zombies).

The current study's data deals with 33 participants from the larger study. While the larger study included gaming groups (i.e., participants given a gaming task), these 33 participants (39.4% reported male gender, 72.7% Caucasian) were all from the mobile multimedia app group, as the focus of this study was on emotions and *learning*. The 33 learners were chosen based on proportional learning gains (also known as normalized learning gains, or relative learning gains). Relative to raw scores, proportional learning gains take both pre-test and post-test scores into account and hence better reflect learning (Taub, Azevedo, Bradbury, Millar, & Lester, 2018). Learners with a pre-test score of 9 and higher (maximum score of 14) were not eligible for the current analyses because this would leave limited variance with which to observe learning to occur. Figure 1 summarizes where the samples originated from.

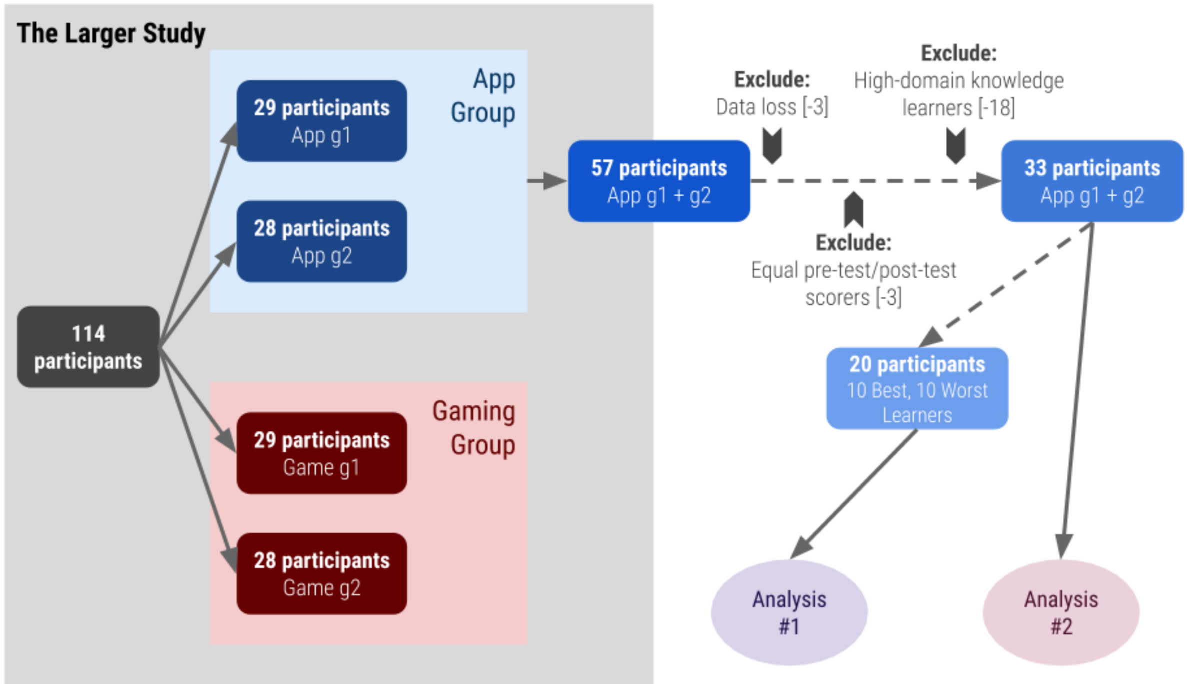


Figure 1. flow chart that summarizes where our samples came from. The "App Group" and "Gaming Group" each has two groups divided between them due to counter balancing the pre-test and post-test knowledge questionnaires. The current study focuses on the 33 participants that were obtained after screening for high-domain knowledge learners from the 57 participants (who were given the learning app) in the larger study.

Multimedia Mobile App

The Edmonton Queer History App (EQH App) features eight locations that showcase the challenges the Edmonton queer community endured, and the social changes that occurred over several decades up to present time. Each location features a multimedia experience, where users are presented with elements such as a video clip, an audio clip, snapshots of artifacts with text captions, a digital maps with a "Street View" function (via Google Maps) and paragraphs of texts. The app contains a tutorial video and a glossary to guide users on using and understanding the app. The app is accessible through any modern internet browser (e.g., Mozilla FireFox) across multiple platforms (e.g., laptop, smartphone).

Measures

Learning gains. Proportional learning gains were measured using a pre-test and post-test knowledge check questionnaire. All except for one question (true or false) were multiple choice questions with four foils. Questions were created by two team members with expertise in LGBTQ+ rights and history. The internal reliability of the 28 questions was low with a Cronbach's Alpha score ($\alpha = 0.55$; Harley et al., 2019), as expected when targeting many different topics within an app (Harley et al., in press; Rowe et al., 2017). The counter-balancing procedure led to the current study having 22 participants taking version A tests, and 11 participants taking version B tests. Despite the different questions between these versions, the topics and difficulty of the test items were balanced through using Papenberg's (2018) minDiff R-package.

Attention. Studies (e.g., Harley, Poitras, Harrell, Duffy, & Lajoie 2016; Knörzer, Brünken, & Park, 2016) have referred to the eye-mind hypothesis (Just & Carpenter, 1980) to infer attention toward learning materials—a perquisite for both learning (Alemdag & Cagiltay, 2018) and emotions (Shuman & Scherer, 2014). Eye gaze data were recorded through EyeLink 1000 and the accompanying software, including a screen-recording feature that captures the screen activity as a video. The video included a blue dot indicating where the learner's gaze was. The EyeLink 1000 was set to Desktop Remote Mode and used the 25mm lens to record the left-eye gaze monocularly, with a sampling rate of 500Hz. The experimenter conducted nine-point calibration/validation for each participant, and further conducted a drift check/correction.

The EyeLink screen-recording indicates a learner's gaze with a blue dot on the screen. To determine whether the learner's gaze was on the content directly related to answering the knowledge-check questionnaire, two processes took place: 1) defining Area of Interests (AoIs), and 2) establishing a rule for determining if the gaze was within the AoIs. Four trained coders were able to identify whether there were gazes on the AoIs based on these two processes. Please see Appendices for detailed explanations of the processes, and the rationale for opting to use human coders.

Emotions. We used FaceReader 7 in this study as our software for automatic facial detection. Noldus, the company behind FaceReader, reported that FaceReader 6.1 had an overall accuracy rating of 95%, with some expressions having higher or lower accuracy (lowest accuracy was 89.1% for angry; highest was 98.5% for surprised; Loijens, Krips, Grieco, van Kuilenburg, den Uyl, & Ivan, 2016). Other studies have investigated FaceReader's quality, with Lewinski, den Uyl, and Butler (2014) showing that FaceReader 6 achieved an 88% accuracy for basic emotions and a 0.69 Facial Action Coding System (FACS) certification score (0.01 point away from receiving the certification). In sum, FaceReader is competitive to human expert coding. Emotions were measured by feeding a recorded face video of the learner to FaceReader 7, which outputted a log file. The log file contained timestamps of when emotions and neutral states were experienced. The discrete emotion "surprised" was dropped from the analysis as manual inspection of the video recordings showed that "surprised" often indicated that the participant was yawning.

Procedures

The larger study, which is where this study's data came from, took place in a laboratory setting, one participant at a time. Each participant completed a demographics form prior to attending. The experimenter first obtained consent from the participant and then gave short instructions on how to interact properly with the measurement equipment (e.g., having proper posture for the eye-tracker). Then, the experimenter briefed the learner on whether they would be interacting with a game or a history learning app. The participants were then given questionnaires. The control group was given a game to play for 30 minutes. The experimental group was given a short tutorial video on how to use the EQH App and were given as long as they needed to complete the virtual tour (*M* = 37.05 minutes, *Median* = 37.72 minutes, minimum = 10.9 minutes, maximum = 55.53 minutes). Learners in both conditions were given the post-interaction questionnaires before they were debriefed and dismissed.

Data Analyses

Data analysis 1. We used 20 of the 33 participants we had in this initial, labor-intensive analysis: The 10 who had the highest degree of proportional learning gains and the 10 who had the lowest. Based on preliminary results from descriptive statistics, we observed that participants lacked emotions during AoI gazes. In other words, the part of the app's content that is directly related to answering the knowledge questionnaire did not seem to trigger facial expressions. This combined with the labor-intensive analysis as illustrated in the appendices, we have decided that analyzing all of the 33 participants data would be outside of the study's scope.

We have focused on the dominant emotions of the participants, which were identified as the emotion that was expressed the most in terms of duration. That is, if a learner expressed sadness for 2 seconds, happy for 5 seconds, and anger for 8 seconds, the dominant emotion for the learner would be anger. We have focused only on the dominant emotions, as opposed to focusing on other emotions or co-existing emotions as our data revealed that the participant's dominant emotion on average accounted for 91.8% of the emotions they experienced (*SD* = 12.5%, Median = 100%). This finding aligns with previous empirical research and hence was unnecessary for us to consider other emotions (Harley et al., 2015; Harley, Bouchet, & Azevedo, 2012).

Data analysis 2. Descriptive statistics were calculated to answer RQ2a and RQ2b, with dominant emotions identified in the same way as the first data analysis. Further, a participant with invalid data, likely due to poor model fit, was dropped. The participant was reportedly scared for 97% of the time, which is a very skewed and unlikely (sustained) emotional response. For RQ3, after capping the score of the outliers to the next most extreme value, and checking for assumptions for a one-way ANOVA, it was revealed that the dataset had an asymmetric distribution. Hence, the Kruskal-Wallis H-test was chosen (a non-parametric alternative to one-way ANOVA; Chan & Walmsley, 1997; Hecke, 2012). In addition to the aforementioned participant being dropped, five other participants' data were also discarded due to them lacking dominant emotions.

Results

Analysis 1

Learners' overall gaze duration toward AoIs ranged from 51.5 seconds to 485 seconds (M = 207 seconds, SD = 92.5). During all of those AoI gaze durations, learners' facial expression (emotion) duration ranged from 0 to 28.9 seconds (M = 2.83 seconds, SD = 6.75). In other words, learners showed facial expressions for 0% to 15.7% of the time when gazing at the AoIs (M = 1.49%, SD = 3.70). As noted by the mean values, the average learner showed little to no emotion when gazing at the AoIs. In fact, when counting the frequencies of the dominant emotions, 65% (13) of the learners had not expressed any emotions at all during their AoI gazes, and hence had no dominant emotion. Anger, scared and sad were each dominant emotions for 10% (2) of the learners respectively, with happy being the dominant emotion for 5% (1) of the learners. When counting all instances of when emotions were expressed (not just the dominant emotions), there was also an instance of contempt that was detected by one learner, expressed for approximately 1.4 seconds. Overall, little emotion was detected during the AoI gazes.

Analysis 2

Frequency, duration and types of emotions during the learning session. Learners' learning session durations ranged from 1,069 seconds to 3,332 seconds (*M* = 2,309 seconds, *SD* = 564). During those learning sessions, learners' total duration of emotions ranged from 0 seconds to 155 seconds (*M* = 25.2 seconds, *SD* = 37.9). In other words, learners showed facial expressions for 0% to 11.1% of the time during gazing at the AoI (*M* = 1.20%, *SD* = 2.15). The second analysis also revealed little emotion, but had a higher absolute frequency of emotions, likely due to capturing expressions during the entire learning session. While only 35% (7/20) of the learners expressed any emotions during the AoI gazes in the first analysis, the second analysis showed that 84.4% (28/32) of learners expressed emotions during the learning session. We note that the additional participants slightly lowered this percentage; the original 20 participants would have shown that 90% of the learners expressed emotions. In terms of the dominant discrete emotions, anger accounted for 18.8% (6), happy accounted for 15.6% (5), sad accounted for 37.5% (12), and scared accounted for 12.5% (4) of the learners. When these emotions are categorized accordingly to their valence and activation, 31.3% (10) of the learners had negative-activating dominant emotions (i.e., anger, scared), while 37.5% (12) had negative deactivating emotion (i.e., sad), and 15.6% (5) had dominant positive-activating emotion (i.e., happy).

Proportional learning gain sores and emotions. We investigated whether there were statistically significant differences in learning gains depending on the expressed dominant emotions. The Kruskal-Wallis H test was conducted to determine if there were statistically significant differences between the “anger” (n = 6), “happy” (n = 5), “sad” (n = 12), and “scared” (n = 4) groups. The distributions of normalized learning gain scores were statistically different between groups, $\chi^2(3) = 8.58$, $p = .035$, $\varepsilon^2 = .33$. The Dwass-Steel-Critchlow-Fligner pairwise comparisons revealed statistically significant difference in proportional learning gains between the anger group and the sad group (W = -4.074, $p = .021$).

Discussion and Significance

Research Question 1a and Research Question 2a

Results indicated that learners did not tend to express many facial expressions while learning with the EQH App, supporting our hypotheses. This was especially the case during AoI gazes. Other studies (e.g., Harley et al., 2015; Jarrell et al., 2016) with learning scenarios comparable to ours reported overall low frequencies of emotions, and hence, we posit our findings are generally aligned with previous research. We acknowledge, however, that learners should experience at least some emotions when dealing with topics that are controversial and sensitive. For example, Trevors, Muis, Pekrun, Sinatra, & Muijselaar (2017) found that reading about conflicting texts regarding climate change elicited emotions in learners, due to cognitive incongruity. We believe the EQH App may elicit similar responses, depending on the learners’ personal experience and beliefs. The low observed frequencies of emotions in our study may be attributed to participants’ expressions underrepresenting the actual incidence of emotions being experienced, as facial expression are just one of the expressive components of emotions and one that may be better at capturing higher intensity emotions. Indeed, the absence of emotions facially, does not preclude the experience of emotions internally or physiologically, especially lower-intensity emotions that may be easier to suppress. Moreover, the CVT would state that social components (human-human interaction), something our learning scenario lacked, may have led to more emotions being expressed via facial expressions (Pekrun & Stephens, 2012).

Research Question 1b and Research Question 2b

Counter to our hypotheses, results revealed that we had a low number of participants that had positive facial expressions, in contrast to previous studies with the EQH App (Harley et al., 2019), and other mobile apps with historical topics (Harley et al., 2016, in press; Poitras et al., 2019). Instead, our results showed a high number of participants with negative dominant emotions such as anger, anxiety/fear or sadness. We propose that the negative emotions observed may be due to the nature of the historical content the app provided, and learners' empathy and emotional capacity to appreciate the historical contexts. The narrative formed through news clips, interviews and pictures may have elicited anger in the learner as some form of outrage toward the mistreatment of queer people. Learners may have further expressed anxiety and sadness as they observed and imagined the violation of human rights queer people experienced. The perspective taking of the challenges the queer community faced may have been more emotionally impactful than the enjoyment of learning experienced in previous studies. In sum, the participants may have felt such emotions on account of being emotionally invested and engaged with the content. High appraisals of task value in a previous study and the current one with the EQH App support this interpretation (Harley et al., 2019).

Research Question 3

Our findings showed that learners' proportional learning gain scores, when split by the learners' dominant emotion, resulted in a statistically significant difference. Specifically, learners who had an angry emotional profile had statistically significantly higher proportional learning gains than those who had a sad emotional profile—counter to our hypothesis and the general prediction the CVT provides. Yet, literature (Pekrun, 2014; Loderer et al., in press) has shown that emotions such as anger, do have the potential to improve learning in certain scenarios. As explained previously, the nature of the content may have elicited anger from certain participants, due to the immersive nature of the EQH app leading to emotional investment. This emotional engagement may have led to cognitive resources being devoted to processing information about the events and locations featured in the app. The CVT does account for relatively less common scenarios such as this and provides some support for our interpretation by stating that emotions such as anger can be beneficial to learning, provided that this emotion is resolved, and promotes engagement. We believe that the EQH App's illustration of how the queer community held its ground and greatly improved social justice in Edmonton may have appeased learners' anger and fostered cognitive processes instead of supporting task irrelevant thinking, which anger tends to.

In the same line of thinking, while sharing the same valence with anger, sadness may have lacked the intensity that led to motivational engagement. Participants with negative-deactivating emotions may have been immersed but thought that social justice was not achieved in the end (lack of appeasement) and may ultimately have lacked the motivation to learn as much as participants who felt anger. Learners with positive-activating emotions (happy) showed high learning performance, as expected by CVT.

The emotion "scared" (anxiety) is also negative and activating, and hence could have aided learning just like anger did. Pekrun & Perry (2014) specifically addresses the potential for this finding, as they stated another dimension may be present to distinguish the two emotions and their functions. Specifically, anger is said to be approach-related, while anxiety is avoidance-related (Carver & Harmon-Jones, 2009). Anger is associated with one exerting themselves towards a situation. If a progress toward a goal is violated by something or someone, anger may be elicited and fuel motivation to directly intervene. Anxiety, however, would promote avoidance away from a situation instead. A person who is angry about mistreatment would want to fight it; a person anxious toward mistreatment would want to avoid it. This difference may lead to the observed difference of the roles these two emotions played in learning. Angry learners may have been more motivated to learn more about the mistreatments and how the events in the EQH App unfolded. Learners who were anxious may have been more motivated to skim and not overly invest themselves in learning what exactly happened.

Limitations

Limitations of this study include those associated with instruments and the sample size. First, the eye-tracker was subject to error and noise. For example, the EyeLink 1000 with the desktop mount required the participants to have a consistent posture, which was sometimes difficult for participants to manage consistently during the learning session. Further, FaceReader 7 is also the subject to the same problem: it output questionable data for one participant, most likely due to an error in calibration or due to the subpar quality of that particular webcam footage. The knowledge test questionnaire also has room for improvement in future and similar eye-tracking studies. Specifically, the “all of the above” questions posed a significant challenge for defining question-related AoIs and had to be removed. Further, due to the large topic to questions ratio, the internal reliability was low.

The generalizability of this study is limited by the size and nature of the sample. Screening for high-domain knowledge learners have resulted in the descriptive and inferential analyses examined a subsample of a study with a small-to-medium sample size. Moreover, inferential analyses were conducted on even smaller samples based on FaceReader’s classification of participants’ emotions and our approach to creating and analyzing dominant emotions as a multi-level independent variable. Inferential analyses and associated results should therefore be treated as preliminary and exploratory in nature, rather than causal. This is especially the case with analysis #1, where we use 10 participants with highest proportional learning gains, and 10 lowest.

Contributions and Future Directions

This study highlighted the incidence of emotions expressed from facial behaviors and the association of such emotions with learning about queer history through a multimedia mobile app. While the LA literature does look at emotions, advanced technologies such as eye-tracking and facial recognition software have been under-utilized. Using them together to explore how emotions can impact learning therefore makes several contributions to the literature, including novel analytical and alignment approaches described in the appendices. Preliminary findings from this study suggest that learning analytics concerning emotions can be complex and nuanced; that typically undesirable emotions such as anger should not be summarily dismissed from instructional design, but rather more closely examined. Moreover, subject domains and varying learning scenarios can have different relationships with emotions. The CVT offers helpful guidelines for predicting achievement-related outcomes, but insufficient detail to account for all learning scenarios. This study provided preliminary evidence that negative-activating emotions such as anger can be beneficial for learning, provided that it promotes motivation and is appeased during the learning process. Feeling anger from historical events may be an indicator for emotional engagement, successful perspective taking, and a deep sense of immersion toward the learning experience—something that may apply very differently in a subject domain such as mathematics.

Future directions include incorporating other data channels such as self-report measures and physiological measures (e.g., EEG) to better measure emotional states. Further, adding additional AoIs that contain elements designed to elicit strong emotional responses such as anger would help extend this line of research by specifically evaluating the explanations we proposed regarding the potential role of anger in social justice history education. Finally, another prominent future direction is the incorporation of emotion regulation as part of LA. While emotions may be elicited, how much, and in what form they are expressed depends on emotion regulation processes (Harley, Pekrun, Taxer, & Gross, 2019). Exploring how learners employ different emotion regulation strategies may help draw a clearer picture of what occurs in a learning scenario with the EQH App and similar apps. Further, being able to recognize, predict, or model emotion regulation strategies, the tendency to use them (or not) and their outcomes on learning in different domains and contexts represent plum opportunities for multimodal LA research programs.

Statements on open data, ethics and conflicts of interest

Any request to access the data should be directed to the second author (Jason M. Harley; jason.harley@mcgill.ca). An IRB approval was obtained prior to the conduct of the study and informed consent was obtained from participants. We declare that we do not have any conflicts of interest regarding the study.

Acknowledgements

This research was supported by funding from the Social Sciences and Humanities Research Council of Canada (SSHRC; Grant ID: 430- 2016-01133), and the Killam Research Foundation awarded to the second author.

References

Carver, C. S., & Harmon-Jones, E. (2009). Anger is an approach-related affect: Evidence and implications. *Psychological Bulletin*, *135*(2), 183–204. https://doi.org/10.1037/a0013965

Chan, Y., & Walmsley, R. P. (1997). Learning and Understanding the Kruskal-Wallis One-Way Analysis-of-Variance-by-Ranks Test for Differences Among Three or More Independent Groups. *Physical Therapy*, *77*(12), 1755–1761. https://doi.org/10.1093/ptj/77.12.1755

Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018). From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning*, *34*(4), 338–349. https://doi.org/10.1111/jcal.12288

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Ekman, P., & Friesen, W. V. (1978). *The Facial Action Coding System*. Palo Alto, CA: Consulting Psychological Press.

Harley, J. M. (2016). Measuring Emotions. In *Emotions, Technology, Design, and Learning* (pp. 89–114). https://doi.org/10.1016/B978-0-12-801856-9.00005-0

Harley, J. M., Bouchet, F., & Azevedo, R. (2013). Aligning and Comparing Data on Emotions Experienced during Learning with MetaTutor. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Artificial Intelligence in Education* (Vol. 7926, pp. 61–70). https://doi.org/10.1007/978-3-642-39112-5_7

Harley, J. M., Bouchet, F., Hussain, M. 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. https://doi.org/10.1016/j.chb.2015.02.013

Harley, J. M., Lajoie, S. P., Tressel, T., & Jarrell, A. (in press). Fostering positive emotions and history knowledge with location-based augmented reality and tour-guide prompts. *Learning and Instruction*, S095947521730405X. https://doi.org/10.1016/j.learninstruc.2018.09.001

Harley, J. M., Liu, Y., Ahn, T. B., Lajoie, S. P., Grace, A. P., Haldane, C., Whittaker, A., McLaughlin, B. (2019). I've got this: Fostering topic and technology-related emotional engagement and queer history knowledge with a mobile app. *Contemporary Educational Psychology*, *59*, 101790. https://doi.org/10.1016/j.cedpsych.2019.101790

Harley, J.M., Pekrun, R., Taxer, J.L., & Gross, J.J. (2019). Emotion regulation in achievement situations: An integrated model. *Educational Psychologist*, 54(2), 106-126.

Harley, J. M., Poitras, E. G., Jarrell, A., Duffy, M. C., & Lajoie, S. P. (2016). Comparing virtual and location-based augmented reality mobile learning: Emotions and learning outcomes. *Educational Technology Research and Development*, *64*(3), 359–388. https://doi.org/10.1007/s11423-015-9420-7

Hecke, T. V. (2012). Power study of anova versus Kruskal-Wallis test. *Journal of Statistics and Management Systems*, *15*(2–3), 241–247. https://doi.org/10.1080/09720510.2012.10701623

Jarrell, A., Harley, J. M., & Lajoie, S. P. (2016). The link between achievement emotions, appraisals, and task performance: Pedagogical considerations for emotions in CBLEs. *Journal of Computers in Education*, *3*(3), 289–307. https://doi.org/10.1007/s40692-016-0064-3

Lewinski, P., den Uyl, T. M., & Butler, C. (2014). Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. *Journal of Neuroscience, Psychology, and*

Economics, *7*(4), 227–236. https://doi.org/10.1037/npe0000028

Loderer, K., Pekrun, R., & Lester, J. C. (2018). Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments. *Learning and Instruction*, S095947521730436X. https://doi.org/10.1016/j.learninstruc.2018.08.002

Loijens, L., Krips, O., Grieco, F., van Kuilenburg, H., den Uyl, M., & Ivan, P. (2016). *Noldus FaceReader: Tool for automatic analysis of facial expression. Reference Manual Version 7.* Publisher location: Wageningen, Noldus Information Technology.

Long, P., & Siemens, G. (2011). *Penetrating the Fog: Analytics in Learning and Education*. 6.

Nistor, N., & Hernández-Garcíac, Á. (2018). What types of data are used in learning analytics? An overview of six cases. *Computers in Human Behavior*, *89*, 335–338. https://doi.org/10.1016/j.chb.2018.07.038

Papenberg, M. (2018). minDiff-Package: The minDiff Package. https://github.com/m-Py/minDiff

Pekrun, R. (2006). The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review*, *18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pekrun, R. (2011). Emotions as Drivers of Learning and Cognitive Development. In R. A. Calvo & S. K. D'Mello (Eds.), *New Perspectives on Affect and Learning Technologies* (pp. 23–39). https://doi.org/10.1007/978-1-4419-9625-1_3

Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement Emotions and Academic Performance: Longitudinal Models of Reciprocal Effects. *Child Development*, *88*(5), 1653–1670. https://doi.org/10.1111/cdev.12704

Pekrun, R., & Perry, R. P. (2014). Control-Value Theory of Achievement Emotions. In

International Handbook of Emotions in Education. https://doi.org/10.4324/9780203148211.ch7

Poitras, E. G., Harley, J. M., & Liu, Y. S. (2019). Achievement emotions with location-based mobile augmented reality: An examination of discourse processes in simulated guided walking tours: Achievement Emotions with Augmented Reality. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.12738

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. https://doi.org/10.1016/j.chb.2017.03.043

Taub, M., Azevedo, R., Bradbury, A. E., Millar, G. C., & Lester, J. (2018). Using sequence mining to reveal the efficiency in scientific reasoning during STEM learning with a game-based learning environment. *Learning and Instruction*, *54*, 93–103. https://doi.org/10.1016/j.learninstruc.2017.08.005

Taub, M., Azevedo, R., Rajendran, R., Cloude, E. B., Biswas, G., & Price, M. J. (in press). How are students' emotions related to the accuracy of cognitive and metacognitive processes during learning with an intelligent tutoring system? *Learning and Instruction*, S0959475218304857. https://doi.org/10.1016/j.learninstruc.2019.04.001

Trevors, G. J., Muis, K. R., Pekrun, R., Sinatra, G. M., & Muijselaar, M. M. L. (2017). Exploring the relations between epistemic beliefs, emotions, and learning from texts. *Contemporary Educational Psychology*, *48*, 116–132. https://doi.org/10.1016/j.cedpsych.2016.10.001

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior*, *89*, 98–110. https://doi.org/10.1016/j.chb.2018.07.027

Appendices

Appendix A

This section outlines the process of establishing areas of interest (AoI), rules for determining whether gaze was present in AoI, and aligning eye-tracking data with facial recognition data.

Defining the Area of Interests

The AoIs were defined by identifying sentences that directly answered the question items from the knowledge-check questionnaire. The sentences were wrapped in a red box for trained coders to easily identify (see Appendix B). Up to 10 questions with the highest incorrect rates were chosen to establish AoIs. Further, questions where the answers contained “All of the above” were dropped due to uncertainty it brought to identifying appropriate gazes.

Gaze Tracking Screen Capture Video

The EyeLink screen capture videos showed where the learner was looking at a certain time but did not necessarily clearly show whether or not the gaze was within the AoIs. To be specific, the EyeLink gaze screen capture video shows a blue dot, indicating where the learner's gaze is on the screen. However, while it is easy to roughly identify where the AoI is on the screen (i.e., someone might say “the blue dot seems close enough to that word, so that should count as the gaze being in the AoI”), it is difficult to be consistent frame by frame. To address this problem, we came up with a systematic way of deciding whether a learner was looking at an AoI. First, we defined “gaze duration” as when the blue dot (indicating the gaze) appeared and continued to show up at least once every one whole second. In other words, if a gaze appeared at 00:00:01.5 (hh:mm:ss.0) and disappeared at 00:00:01.8, but appeared again before a whole 1-second elapsed (from 00:00:01.8, which would be 00:00:02.8), the gaze instance would be considered to be still active. If the gaze disappeared at 00:00:01.8, and appeared again after a whole second had passed (e.g., 00:00:03.8), then the time between 00:00:01.5 and 00:00:01.8 would be considered as one gaze instance, while the timestamp 00:00:03.8 would be a starting point for another gaze instance.

The minimum length between gaze instances was chosen to be one whole second due to the limitation of the methodology, specifically regarding the alignment of the gaze data with the facial expression data. If we were solely interested in finding out when a facial expression occurred, we could simply consult the log files FaceReader outputs. However, to figure out when a facial expression occurred in relation to gaze behavior, we needed to do the following: 1) consult the screen capture video, and figure out when a specific gaze instance occurs—we should have a timestamp for when a gaze instance begins, and another timestamp for when a gaze instance ends. 2) For each of the timestamps, the screen capture video is able to tell us the current time (the time of day, not the time elapsed in the video) by looking at Windows 7's system clock on the bottom right corner—we can now convert the timestamps into what the time of day was (i.e., instead of saying a gaze instance began at 0:03:01 of the screen capture video, we can now say it begins at 1:40:04 PM). 3) From the new time stamps obtained, we can now figure out when a facial expression occurred inside the webcam video (1:40:04PM to 0:02:57 of webcam video), as the webcam video will also show the system time at every frame. 4) From here, we can figure out exactly when the facial expression occurred within the video in relation to gaze behavior by consulting the FaceReader state log file.

The limitation primarily comes from step 2 of the process described above; the Windows 7 system clock by default only shows hours and minutes, and with additional tweaks we were able to make it show seconds, but not milliseconds. Because the smallest unit of time we can refer to are seconds, this poses a limitation regarding how granular we could get when dealing with videos. The gaze instances have to be minimally one whole second apart to ensure that the video playback can differentiate each gaze instance by at least one second. In other words, if the threshold for distinguishing the gaze instances was less than one second, there is potential for the webcam video to unintentionally combine two or more multiple gaze instances into one.

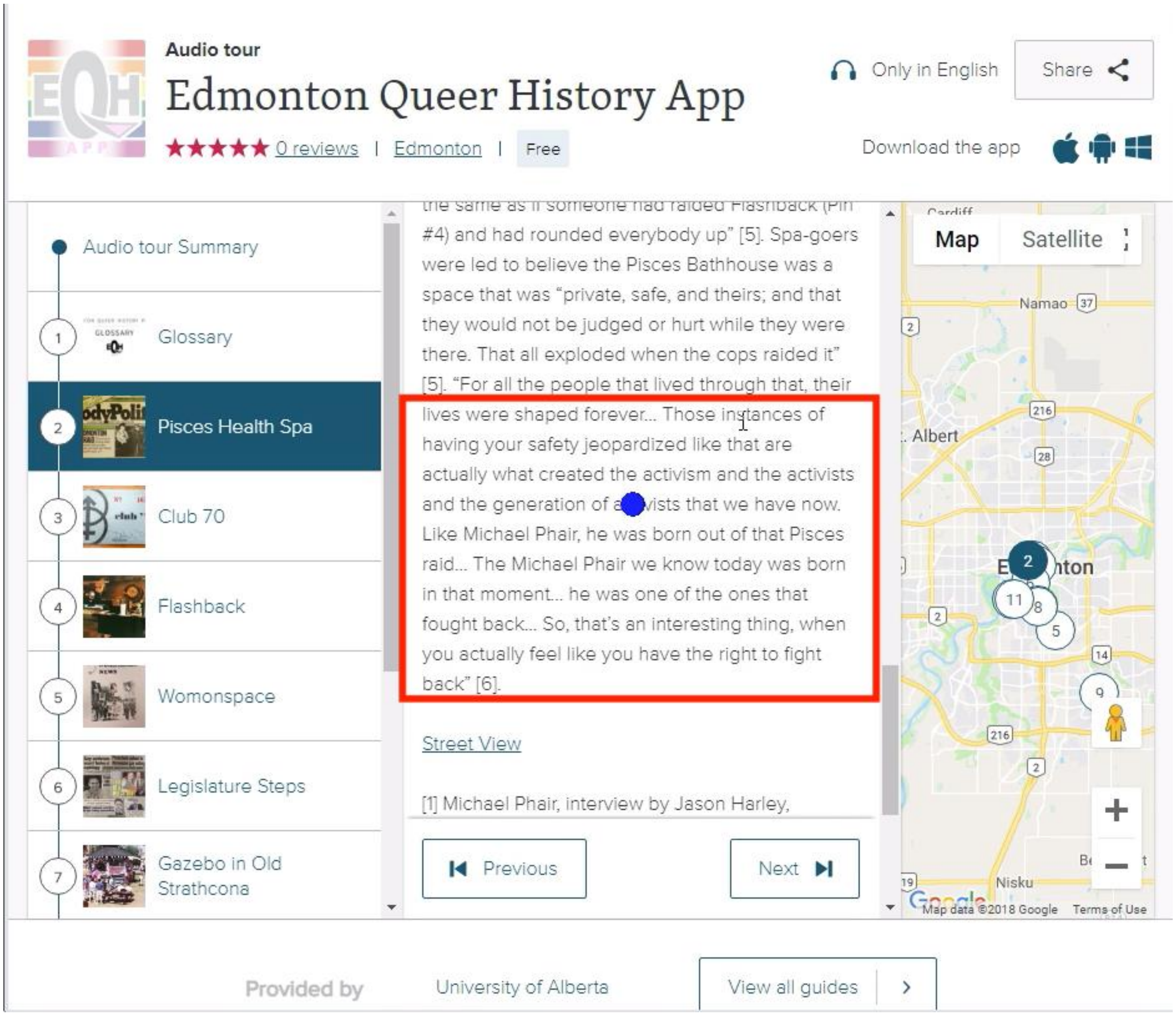
The second part of coming up with a systematic way to determine AoI gazes was to add a grid to the gaze capture videos. Appendix B shows a screenshot illustrating how adding grids to the gaze screen capture video helps with objectivity towards determining AoI gaze. The grid had 27 row cells and 45 column cells. The AoI was predetermined by a red box—this was shown to coders with a screenshot. Appendix C shows an example of this. The coder would refer to the AoI screen capture, then look at the current frame that showed the learner's gaze. If the gaze was within the grid that covered the AoI box, it was considered to be inside the AoI. There are two critical details to this process: First, due to the inaccuracies of eye-tracking caused by uncontrollable variables (e.g., shifts of learners' posture, learners wearing mascara and glasses, etc.) the coders were instructed to count the gaze as an AoI gaze even if the blue dot was one cell off from the AoI box (including one cell off diagonally). Second, if even a single pixel of the blue dot was inside the AoI box or one cell away from the AoI box, the gaze was considered to be an AoI gaze. This rule is helpful in making sure the verifying AoI gaze is simple and consistent. Otherwise, the coders would have had to determine exactly how much of the blue dot had to be inside the AoI to count as a gaze, which would have likely led to disagreements or inconsistent coding.

The gaze instance was recorded by going through the screen capture video frame by frame. Most of the videos were 15 frames/second, while a few had 9 frames/second. In rare instances, the video had some other frame rate, but were adjusted to have 9 or 15 frames/second as long as the original frames/second was higher than 9 or 15, respectively. Viewing the videos frame by frame was important to make sure each gaze instance was recorded properly. Four research assistants were trained in this task, and they coded each learners' videos under direct supervision. The assistants first recorded each gaze instance from the screen capture videos. Then, they identified what timestamp had to be taken from the FaceReader log. Finally, they recorded the start and end of a gaze duration on a spreadsheet.

By consulting the spreadsheet containing all the instances of gaze during the learning session, it was possible to detect when and for how long learners were paying attention to materials relevant to the knowledge check questionnaires. It should be noted that, due to the counter-balancing of the knowledge-check questionnaires, not all learners viewed the same exact questions. That is, one group of learners would have viewed one set of questions (questions #1 to #14), and the other group would have viewed the remaining set (questions #15 to #28). Not all of the questions were analyzed, but instead the 10 questions with the lowest correct-answer rates from the learners were chosen. For example, the question "In its early years, what was the main purpose of the first Pride parades?", had the lowest correct-answer rate from the leaners, with only 34% of the learners answering the question correctly. Further questions were dropped if the questions contained more than one possible AoI that could have led to the learner answering the question correctly. For example, questions that had the answer choice "D) all of the above" implies that the learner could solve the question without investing eye-gaze towards reading material related to answer choices A) to C). Instead, it is reasonable to imagine that the learner could deduce the answer if they had eye gaze towards reading content related to just two of those choices. Due to the complications this brought, one group of learners (n = 8) had eight questions analyzed as opposed to ten.

Appendix B

Screenshot that tells the coders where the AoI is. By referring to this screenshot and the gaze capture video with the added grids, it is possible to determine systematically whether or not the blue dot is inside the AoI.



Appendix C

Screenshot of the EyeLink gaze screen capture video with grids added on. The added grid can help determine whether the gaze (blue dot) is inside the AoI or not.