Learning About Queer History with a Multimedia Mobile App: The Role of Emotions

by

Byunghoon Ahn

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Education

in

Technology in Education

Department of Educational Psychology

University of Alberta

Abstract

Achievement emotions are important in learning, as they influence learners' cognitive, motivational, and regulatory processes. Pekrun's (2006) Control Value Theory (CVT) of Achievement Emotions associates positive-activating emotions (e.g., enjoyment) with better learning outcomes and the opposite for negative-deactivating emotions (e.g., hopelessness). Other emotions lead to mixed results, but emotions such as anger, on average, show loss in learning performance. This thesis investigated what emotions were generated as learners interacted with a multimedia mobile app to learn about queer history in a North American city: Edmonton. I further examined whether learners' emotional profiles were associated with different levels of learning outcomes. Emotions were measured with an automatic facial recognition software (FaceReader 7). Results indicated that learners tended to show little emotion, facially. An investigation of dominant emotion profiles revealed that learners expressed more negative-activating emotions (anger, anxiety), and negative-deactivating emotions (sadness) than positive-activating emotions (happiness). Learners that expressed anger as their dominant emotion, had the highest learning performance; one that was statistically significantly different from learners with a sad dominant emotion profile. This study adds to the field of emotions and technology-rich environments (TREs) by integrating a framework for learning-related emotions and utilizing an under-examined emotion measurement methodology. This study further highlights the importance of examining typically undesirable emotions in learning, such as anger, in subject domains such as history.

Preface

This thesis is an original work by Byunghoon Ahn. This research project received research ethics approval from the University of Alberta Research Ethics Board, Project Name: “Fostering Historical Reasoning, Hope, Empathy, Emotional Engagement and Queer History Awareness with a Mobile Augmented Reality App: Part 3”, No. Pro00084257, August, 22nd, 2018.

Acknowledgments

I would like to thank my supervisor Dr. Jason Harley for their time and dedication guiding me throughout the many stages of this thesis. Frodo had Gandalf, Harry had Dumbledore, and Neo had Morpheus; I had Dr. Harley. I would also like to thank Dr. Yang Liu for bringing his technical talent and expertise to the running of experiments, helping to organize the data, and supporting quality checks along the way. I would like to thank members of the Edmonton queer community and CHI-TEA lab for their contributions to making the Edmonton Queer History App project so successful, which in turn, helped make my study possible. I would like to express special thanks towards the undergraduate CHI-TEA lab members that took on the labor-intensive tasks of organizing and cleaning the data: Isabelle Estrada, Darby Gynane, YeWon (Annie) Han, and SoWon Kim. My family and friends always deserve my thanks for their unconditional support and love. I also thank Mihee Park for being my mental stabilizer by providing unending waves of emotional support. I thank Ellen Kim for believing in me. This research was made possible by funding awarded to Dr. Harley (PI) from the Killam Research Foundation and funding awarded to Dr. Harley (PI), Dr. Lajoie (co-PI), and Dr. Grace (co-PI) from the Social Sciences and Humanities Research Council (SSHRC) of Canada.

Table of Contents

List of Tables

List of Figures

Introduction

Attention towards emotions in education has come a long way: studies prior to the 21st century have tended to focus solely on students' test anxiety and neglect other emotions (Pekrun, Hall, Goetz & Perry, 2014; Schutz & Pekrun, 2007). The past couple of decades have seen an exciting shift, however, with increasing recognition of the crucial roles other emotions play in education (Linnenbrink-Garcia & Pekrun, 2011; Pekrun & Stephens, 2012). Pekrun and Stephens report that students often experience emotions other than anxiety that can be classified as positive emotions (e.g., enjoyment, hope, pride, relief) and negative emotions (e.g., anger, shame, boredom). Both positive and negative emotions have been examined extensively in relation to many educational topics such as cognitive and behavioral engagement (e.g., Reschly, Huebner, Appleton, & Antaramian, 2008; King & Gaerlan, 2014), problem-solving (e.g., Clore & Huntsinger, 2007; Spering, Wagener, & Funke, 2005) and self-regulated learning (e.g., Pekrun, Goetz, Titz, & Perry, 2002a). Overall, the emerging research has shown that emotions influence one's cognitive resources, motivation, use of strategies, and self-regulation (Pekrun, Goetz, Titz, & Perry, 2002b; Pekrun, Frenzel, Goetz, & Perry, 2007).

This thesis set out to examine learners' emotional experiences while they used a multimedia mobile app to learn about queer history. I did so by examining the emotions experienced by learners in real-time using automatic facial recognition software, including what kinds of emotions they expressed while (a) they read content directly related to answering a knowledge-test questionnaire and (b) across the whole learning session. I then examined whether expressing certain emotions during (a) and (b) influenced learners' knowledge outcomes. My research will hence contribute to exploring the impact certain emotions can have when it comes to learning outcomes. I have used Pekrun's (2006; Pekrun & Perry, 2014) control-value theory

(CVT) of achievement emotions to inform my investigation of emotions and learning, including the generation of research questions and hypotheses. Further, I have referred to Ekman's (1992) theory of basic emotions to support my inferences regarding learners' discrete emotions captured using automatic facial recognition software.

Theories and Frameworks of Emotion

This section describes the definition of emotion, and the supporting theoretical frameworks I have chosen for this study. These choices have shaped my research questions and the methodology I have used to answer them.

Defining emotion. For this thesis I have defined emotions as multi-componential psychological responses, that are generated towards object foci of situations perceived to be relevant to one's goals (Gross 2010; Harley, Pekrun, Taxer, & Gross, 2019). The multi-componential aspect of emotion features an orchestra of affective, cognitive, physiological, motivation, and expressive processes (Pekrun, 2011). For example, imagine a student frantically studying for an exam the day prior to it. The focus of the object may be her progress relative to how much time she has left. She is constantly cognizant of the current time as she focuses on how long it has taken her to get to just chapter four: six hours to get this far, with the current time being 11:00pm. She might constantly look at how many pages she has left to go: 238 pages. The constant focus on stimulus leads to evaluation of her situation and goals, and ultimately outputs a feeling of nervousness (affective process), imagination of worst-case scenarios (cognitive process), sweaty palms (physiological process), determination to stay up late (motivation process), and shaking of legs and biting lips (expressive process). The aggregate of these components is what makes an emotion.

This definition is helpful in a few ways due to the cascades of implications it brings. First, that emotions are event-focused (Scherer, 2005); a response to something. Without assigning attention toward a stimulus first, emotions will not be generated. Further, by perceiving the stimuli and assessing it accordingly to one's goals, the process of emotion generation is appraisal based and goal-driven. These implications help distinguish emotions from other affective factors. If emotions are an event-focused response, the evaluation towards that event in generating a response will take place constantly. Each iteration of this evaluation can lead to differing emotions, and hence emotions will tend to be short (Shuman & Scherer, 2014). Moods, on the other hand, lack object of focus, and hence linger longer than emotions (Gross & Thompson, 2007; Gross, 2010; Pekrun, 2011). Attitudes last even longer than moods, and while having similar components to emotions on the surface, they differ considerable when it comes to cognitive facets (attitudes is on beliefs vs. emotions is on real-time perception); structure of components (emotions have more complex structures in both affective and behavioral components); and behavior relation (attitudes have weaker relation to behavior; Shuman & Scherer, 2014).

Other implications include the definition being compatible with the mainstream view of what emotions are, as definitions of emotions across different theories tend to accept emotions as componential episodes (Shuman & Scherer, 2014). The definition is also aligned with Pekrun's CVT, and other related theoretical models, including Harley and colleague's (2019) Integrated Model of Emotion Regulation in Achievement Situations (ERAS), which is derived from and extends both Gross' and Pekrun's theories. Ultimately, by utilizing a definition that is widely accepted and backed up by prominent theoretical frameworks, this thesis can bring work that is generalizable and digestible across many different viewpoints.

The Control-Value Theory of Achievement Emotions. CVT examines achievement emotions — emotions related to achievement activities or outcomes (Pekrun, 2006). Emotions elicited from viewing a lecture video on a laptop (i.e., activity emotion), or from viewing and reflecting on the final grade for a course (i.e., outcome emotions) would both be considered achievement emotions. These emotions are explained with two variables. The first is the level of control one feels, and the second is the level of value one feels toward an achievement task. For example, imagine a student is studying for a test. She feels that the test is incredibly important but also feels that she has little chance of doing well, regardless of how hard she studies (low control, high value). In this case, she may feel hopelessness. The level of control and value is determined through cognitive appraisals, and act as antecedents of emotions. CVT describes the elicited emotions with a three-dimensional taxonomy: valence (negative or positive—intrinsic feeling of bad or good), activation (deactivating or activating — also known as arousal), and object focus (retrospective outcome, activity, or prospective outcome). For example, emotions such as happiness or enjoyment during learning would have a positive valence, high activation (activating), and a focus on the activity at hand. In general, high control appraisals are associated with positive emotions, and vice versa (Pekrun, 2011). High perception of value is associated with more intense emotions, and vice versa.

The differing combinations of value appraisals and control lead to different discrete emotions depending on the object focus: focusing on outcomes in hindsight leads to retrospective outcome emotions; focusing on outcomes that are yet to be leads to prospective outcome emotions; and focusing on the current activity leads to activity emotions. For example, a student might have a big presentation coming up, and feels hope as she knows she has practiced numerous times (positive value, medium control, prospective outcome). When the student gives her presentation, perhaps she feels a little bit of enjoyment as she feels her presentation is going smoothly without any mistakes (positive value, high control, concurrent activity). However, when the student gets her grade, she realizes she's gotten a measly D. As she is told her presentation was good, but her content was completely off-topic — she feels anger because she believes her grade was unfair based on the assignment description she read in the syllabus (negative value, control assigned to others, retrospective outcome). This thesis focuses on a range of emotions, including those with positive and negative valence and high and low activation levels that were elicited from an activity (i.e., concurrent activity emotions).

Achievement emotions and learning. CVT purports that emotions influence learning outcomes through their impact on cognitive, motivational, and regulatory processes (Pekrun, 2006; Pekrun & Perry, 2014). Specifically, CVT proposes that because emotion generation requires attention, it likely leads to activation of cognitive resources such as working memory resources. Certain positive emotions such as enjoyment of learning may facilitate devoting these cognitive resources to the achievement activity at hand. On the other hand, emotions such as anxiety or hopelessness will likely inhibit cognitive processes such as attention and memory processes from being focused on learning. Pekrun (2006) specifically points out that enjoyment is likely an exception to emotions helping with cognitive resource allocation rather than the norm; other emotions even with positive valence can take up precious cognitive resources that should be spent on learning.

Further, emotions can impact learners' interest and motivation to achieve (Pekrun, 2006; Pekrun & Perry, 2014). Positive activating emotions can support both intrinsic (e.g., enjoyment in learning and improving one's knowledge) and extrinsic motivation (e.g., pride in getting on the dean's list with a high GPA). Deactivating negative emotions can lead to the perception of low-control (e.g., hopelessness) or low-value (e.g., boredom) towards achievement, hence leading to low motivation and interest. Deactivating positive emotions (e.g., relaxation) and activating negative emotions (e.g., anger) are expected to have more complex relationships with motivation. CVT proposes that deactivating positive emotions may cause immediate disengagement towards learning but may lead to long-term engagement and commitment. For example, relief from getting a perfect score on a mid-term may lead to a momentary drop in motivation to study, but the positive feeling from an excellent result can be a reinforcement towards continuing with the study for the final exam — this would be the opposite of getting a flat out F on a midterm and feeling hopelessness, which may lead to complete and utter disengagement from future studies. Activating negative emotions such as anxiety may lessen intrinsic motivation, but also have the potential to bolster extrinsic motivation of avoiding failure. Despite this potential, CVT would expect that on average, negative activating emotions should lead to loss in learning performance (Pekrun, Lichtenfeld, Marsh, Murayama, & Goetz, 2017).

Emotions can also impact learning strategies and information processing (Pekrun 2006; Pekrun & Perry, 2014). Positive emotions are expected to support top-down, relational, and heuristic processes, with positive activating emotions especially supporting flexible and creative thinking. Learners with positive emotions may employ strategies such as organization of learning material (e.g., mind-mapping). Negative emotions will tend to support bottom-up, item-specific, analytical, and rigid processes, and may ultimately lead to strategies such as rote memorization and simple rehearsal. Negative deactivating emotions such as boredom and anxiety may lead to inhibition of employing any effective learning strategies, and result in mere superficial process of information (e.g., just reading, not even memorizing, the words on a cue-card).

Finally, positive emotions (e.g., enjoyment) are thought to facilitate self-regulated learning, while negative emotions (e.g., anxiety) are thought to promote dependency on external regulation. This is explained through self-regulation strategies requiring flexible cognitive processes, something positive emotions can facilitate as explained above. Thus, general happiness from learning should lead to proactive goal setting, planning, monitoring and evaluation of one's educational experience by the learners themselves. Feelings of hopelessness or frustration may lead one to seek out these helpful processes using external resources (e.g., a tutor) instead. Generally, positive activating emotions are expected to have a stronger impact on self-regulated learning than positive deactivating emotions.

In sum, positive emotions, especially positive activating emotions should lead to better cognitive processing on an achievement activity, overall better utilization of regulation and learning strategies, and higher commitment and engagement towards learning, relative to negative emotions. As explained above, CVT accounts for the complex interplay of these different mechanisms and notes that positive emotions cannot guarantee positive learning results, nor would negative emotions absolutely spell disaster for learning. The tendencies the CVT predicts have been supported by empirical studies (Pekrun et al., 2017; Pekrun & Perry, 2014) and provide guidelines for measuring and understanding emotions in achievement scenarios.

CVT and studies on emotions. The CVT has been implemented in numerous studies that examine learning and emotions (Butz et al., 2015; Harley, Carter, et al., 2016; Harley, Poitras, et a., 2016; Harley, Lajoie, Tressel, & Jarrell, in press; Harley, Liu, Ahn, in press; Pekrun, Goetz, Daniels, Stupinsky, & Perry, 2010), including those in technology-rich environments (TREs) which are of particular relevance to my thesis (others have used similar terms, such as technology based learning environments, TBLE; or computer-based learning environments, CBLE). Loderer, Pekrun, and Lester (in press) have recently conducted a systematic review and meta-analysis on emotions in TREs. TREs in their study included environments and features such as virtual simulation (e.g., Noteborn, Carbonell, Dailey-Hebert, & Gijselaers, 2012), virtual pedagogical agents (e.g., Liew, Tan, & Jayothisa, 2013; Johnson, Rickel, & Lester, 2000; Johnson & Lester, 2016), augmented reality (e.g., Wojciechowski & Cellary, 2013), and others. Results showed that across 149 primary studies, enjoyment was positively correlated with learning outcomes, with a mean correlation coefficient value of .18. This finding was backed up by enjoyment having a moderately strong correlation with engagement and effective strategy use ($\bar{r}$ = 0.45 and 0.31 respectively). Anxiety was negatively correlated with learning outcomes ($\bar{r}$ = -0.14), engagement ($\bar{r}$ = -0.12) and strategy use ($\bar{r}$ = -0.17). As the CVT predicted, anger/frustration, confusion, and boredom presented more complicated relationships. Each reported a mean correlation value of -0.07, -0.09 and -0.08 towards outcomes respectively, but no significant results were found. However, it should be noted that these results are aligned with the CVT: although there will be cases of negative emotions having no negative impact, or even improving academic performance, negative emotions on average should lead to diminished performance. The meta-analysis concluded that promoting positive activating emotions such as enjoyment and reducing negative activating emotions such as anxiety should be beneficial to learning. However, certain negative emotions such as boredom and anger can also be beneficial, under the condition that they promote deeper engagement, and the emotions can be successfully resolved. It should be understood that scenarios where emotions such as boredom aiding learning should be rare (Pekrun et al., 2014), but CVT accounts for exceptions. The findings examined in Loderer and others' study aligns with other studies conducted in non-technology-based environments, summarized in Pekrun's (2014) paper.

Call for CVT: examining emotions in multimedia learning. Despite the numerous studies featured in Loderer and colleagues' (in press) meta-analysis, there are studies that have examined emotions and learning without a framework such as the CVT. This has led to calls for integrating the CVT into different fields of research, such as medical education (Artino, Holmboe, & Durning, 2012a, 2012b) and multimedia learning research (Stark, Malkmus, Stark, Brümlem, & Park, 2018; Um, Plass, Hayward, & Homer, 2012).

The motivation for these calls is rooted in studies tending to examine learning-related emotions without tailored theoretical frameworks. 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 limitation was addressed by Pekrun (2014) where he stated that "it would be inadequate to simply equate positive emotions with positive achievement effects and negative emotions with negative effects" (p.135). Distinguishing academic emotions from general emotions and adding additional layers to describe and explain emotions were part of what made the CVT the ideal lens for this thesis.

The theory of basic emotions. The theory of basic emotions (also known as discrete emotions theory; Ekman, 1992; Ekman & Friesen, 1978; Shuman & Scherer, 2014) takes an evolutionary perspective and has its roots from Darwin's (1998) study on emotion and evolution. It purports that because the neurobiological faculties responsible for emotions have evolved adaptively, emotions are also a product of species adapting for survival (Izard, 2007). For example, feeling intense emotions that may encourage one to escape predators would have contributed to better chances of survival. In other words, maladaptive emotions, or the lack of emotions altogether would be detrimental to survival.

Because these emotions are part of evolution that has shaped *Homo sapiens* as a species, emotions are thought to be universal. Ekman (1992) has identified six such universal emotions: anger, fear, disgust, sadness, happiness, and surprise. These emotions can be further categorized into negative valence (anger, fear, disgust, sadness), positive valence (happiness), and non-valenced (surprise). Surprise is considered to be non-valenced, or neutral, due to it being prone to leaning toward either direction (Harley et al., 2015). The basic emotion theory explains other emotions by considering them as a variation of the six basic emotions, or a mixture of the basic emotions (Ekman 1992). To clarify what variation refers to, it should be noted that Ekman talks about families (i.e., categories) of emotions, where many different forms of responses (e.g., facial expressions, physiological responses) can all point to a single family of discrete emotions. The six basic emotions each form a family. Ekman provides an example of this with anger, where the numerous different “angry” facial expressions share the feature of having eye brows being lowered and drawn together, in addition to the upper eyelids being raised and the lips being tightened. These shared features can take all the different “variations” of anger, and these variations are considered to be a member of the “anger family”. For example, the feeling of frustration, an emotion not directly listed by Ekman, will most likely belong to this “anger family”. It should be noted that the expressions that belong to the anger family do not belong to seemingly related families such as disgust as each family has distinct features. In other words, Ekman’s theory proposes that the boundaries between the discrete emotions are clear.

Basic emotion theory also distinguishes emotion from other types of affect. Ekman proposes nine characteristics that do so: 1) distinctive universal signals (e.g., facial expressions universally corresponding to discrete emotions); 2) presence in other primates; 3) distinctive physiology (e.g., specific autonomic nervous system activity for each discrete emotions); 4) universal antecedent events (e.g., tendencies for specific emotions to be elicited by corresponding stimulus); 5) coherence in response systems (e.g., emotions lead to coordinated responses from different processes such as expressive and autonomic processes); 6) quick onset (i.e., tendency for emotions to begin quickly); 7) brief duration; 8) automatic appraisal (e.g., cognitive processes related to emotions can be executed at a low (subconscious) level); and finally, 9) unbidden occurrence (i.e., people tend to feel emotions, as opposed to choosing to feel specific emotions). These characteristics align with the CVT, and hence is compatible with the definition of emotions I have chosen for this thesis. Specifically, both theories put emphasis on the antecedents of emotions and their impact, while also agreeing that emotions tend to occur in short bursts, and hence can quickly shift in quality. The high degree of compatibility between the theoretical frameworks I have chosen lets me use them together to explore emotions and learning in my thesis. Further, the work on theory of basic emotion has led to the publication of Facial Action Coding System (FACS; Ekman & Friesen, 1978). Action units are muscle groups responsible for facial expressions, and FACS in combination with the theory of basic emotions have informed research on facial expression recognition for studying emotions — the method I use for measuring emotions in this thesis.

Measuring Emotions: Facial Recognition

Harley (2016) reviews how researchers measure emotions when studying emotions in computer-based learning environments (CBLEs). Although self-report measures are the most commonly used for studying emotions, Harley further outlines four additional types of data: facial expressions, body posture, physiological patterns, and log-files. Methods that employ multiple channels are also possible. All of these approaches have pros and cons, and I have chosen to use facial recognition for several reasons. First, facial recognition is based on Ekman and Friesen's (1978) FACS, and hence aligns well with the theoretical frameworks I have chosen. FACS is a system that examines the small groups of muscles that make up a facial expression to determine if the expression falls into one of six discrete emotion categories described by Ekman. Second, facial recognition has higher reliability compared to most other methods (Harley, 2016). Third, facial recognition can be both unobtrusive and detect emotions in real time as opposed to measures such as self-reports. While facial recognition traditionally requires expert human coders, and hence demands a large investment of resources, the introduction of automatic facial recognition software shows promising results, with comparable accuracy to human coders (Harley, 2016) and high correlations with self-reported emotions (Harley et al., 2015). The project this thesis is based on has used FaceReader 7, developed by Noldus. More information on FaceReader is described in the methods section.

Eye-tracking and Attention

While eye-tracking has been used to predict emotions such as curiosity and boredom (e.g., Jaques, Conati, Harley, & Azevedo, 2014), by providing various forms of data such as blinks and pupil sizes (Smilek, Carriere, & Cheyne, 2010; Wang, Chignell, & Ishizuka, 2006), this thesis utilizes eye-tracking to measure attention. Specifically, it measures gaze duration to infer attention towards a stimulus, a valid approach according to the eye-mind hypothesis (Just & Carpenter, 1980). Studies (e.g., Harley, Poitras, Harrell, Duffy, & Lajoie 2016; Knörzer, Brünken, & Park, 2016) have referred to the eye-mind hypothesis to employ methods such as analyzing the percentage of recorded fixation time towards the learning material, and the absolute number of fixations to determine the overall level of attention learners invested in learning. The addition of eye-tracking to measure attention helps, as it is one way of addressing the problem of lacking objective methods of determining whether or not there are cognitive activities — which are prerequisite for both learning (Alemdag & Cagiltay, 2018) and emotions (Shuman & Scherer, 2014) in learning situations where visuals are important.

Related Studies

This study built off of prior empirical research that used automatic facial recognition software to examine emotions during learning or examined emotions during learning with mobile multimedia apps. This study is the first to use automatic facial expression software to examine emotions during learning with mobile apps.

Harley, Bouchet, Hussain, Azevedo, and Calvo's (2015) study featured an intelligent hypermedia learning environment (MetaTutor) that learners used to learn about the human circulatory system. They evaluated the synchronous use of three data channels for measuring emotions: facial recognition technology, self-reports, and electrodermal activity measurements through a wearable bracelet device. The study reported that while facial recognition technology and self-report data had a high agreement rate of 75.6%, lower levels were found between self-report and electrodermal activity (41.3%), and facial recognition and electrodermal activity (60.1%). One of the conclusions this study made was that their results strengthened the validity of utilizing automatic facial recognition software to measure emotions during learning activities. One notable detail of their data analyses was that accounting for secondary co-occurring emotions (as opposed to focusing on one dominant emotion) was not necessary, as the vast majority of the time emotions were examined (80.7%), there were no signs of other dominant emotions for more than 3 seconds. Further, 17.93% of the time were cases when neutral was either the dominant or secondary dominant emotion. In other words, it would be reasonable to presume that learners will experience one strong emotion during learning moments.

Taub et al (in press) have also used MetaTutor to investigate the relationships between emotions from facial recognition software, cognitive and regulatory processes, and learning outcomes. Specifically, they recorded learner activities such as typing out notes, or the interaction with the pedagogical agent to assess learners' learning strategies (taking notes and summarizing), judgments of learning, and feelings of knowing. In addition, they examined participants' facial expression through video recordings. They then examined these factors alongside with learners' proportional learning gains. 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 explained it by suggesting that learners who experienced frustration, were perhaps, better engaged and motivated to take the most informative and accurate notes as the frustration's source was from not fully understanding the learning material. The authors note the helpful potential of negative emotions such as anger and frustration should be further explored.

Another relevant study comes from Jarrell, Harley, and Lajoie's (2016) examination of achievement emotions in BioWorld — a software program to support medical diagnostic reasoning). Using a k-means cluster analysis, 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 study explained that 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 emotion reported 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 explain that perhaps the experience of emotions in cognitive resources inhibited cognitive resourced being devoted to learning.

Harley, Liu, Ahn…et al. (in press) used the same app used in my thesis, although we asked different research questions and used a different sample from an earlier study. Our results included learners reporting high mean values of enjoyment, and low mean values of boredom. Furthermore, Harley, Poitras et al. (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, Lajoie, et al., 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 were able to be clustered into learners with positive and negative emotional profiles, and learners who experienced enjoyment in learning with the app were more successful at identifying distractor information during measurements of topic understanding.

The above studies, while different, highlight common grounds that helped me formulate my hypotheses. Specifically, the studies show 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. Finally, the studies make a case for facial recognition software being a valid method for measuring emotions in real time.

The Current Study: Research Questions and Objectives

This thesis examines the types of emotions generated while participants learned about queer history content with a multimedia mobile app. The thesis further examined if learners with different dominant emotions (i.e., emotions they experienced the most) had significantly different levels of proportional learning gains. This thesis extends the research being carried out in the field of emotions in TREs. It answers the call for integrating theories that are apt at handling learning-related emotions and employs methodologies that have been validated as measures of attention (i.e., eye-tracking) and emotion (i.e., facial recognition software:

FaceReader 7) during a learning activity. To carry out the investigation of emotions in this study, I have proposed the following research questions examined over two separate analyses.

Analyses 1. RQ1a) How often did learners express an emotion while they read content directly related to the answers on the knowledge check questionnaire? b) What kinds of emotions did learners express when categorized by (i) discrete emotions and (ii) valence-activation dimensions? RQ2 Were there statistically significant group differences in proportional learning gains when learners were grouped by their dominant emotion (e.g., anger vs. happiness)?

Analyses 2. Further, these research questions were repeated for the entire learning sessions. To clarify, RQ1a,b and RQ2 in Analyses 1 concerned data only when the learners were gazing at content in the Edmonton Queer History App related to questions on the post learning session knowledge test. The second round of questions in Analyses 2 adjusted RQ1a) so that I examined learners' emotions during the entirety of their learning session, and not just when they were gazing at test item-related content. The same treatment was applied to RQ2. Therefore, the research questions in Analyses 2 were the following: RQ1a) How often did learners express an emotion during the entirety of their learning session. b) What kinds of emotions did learners express when categorized by (i) discrete emotions and (ii) valence-activation dimensions? RQ2 Were there statistically significant group differences in proportional learning gains when learners were grouped by their dominant emotion (e.g., anger vs. happiness)?

Hypotheses. The CVT and previous research that examined emotions with mobile apps or used facial recognition software to examine emotions with mobile apps played a particularly important role in forming hypotheses for the above research questions. Harley and colleagues (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) and others have corroborated this finding that many learners tend to express low levels of emotion. I have therefore hypothesized for RQ1-a (for both sets of analyses) that I would expect learners to express neutral facial expressions most of the time and show facial expressions only once in a while.

Previous findings with the Edmonton Queer History App used in this thesis as well as other similarly designed apps for learning history (Harley, Liu, Ahn, et al., in press; Harley, Poitras, et al., 2016, Harley, Lajoie, et al., in press) 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, I have hypothesized that learners would tend to express discrete emotions of happiness (positive-activating), and tend to express little sadness (negative-deactivating), or anger (negative-activating) for the RQ1-b (for both sets of analyses).

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 2014; Loderer et al., in press). Negative-deactivating emotions such as hopelessness (likely expressed as sad in terms of facial expression), on the other hand, will have a negative impact on learning outcomes. 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 I have hypothesized that learners with positive-activating dominant emotions (i.e., happy) would have higher proportional learning gain scores compared to learners with other emotions. I 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. This hypothesis was for RQ2 for both sets of analyses.

Methods

Participants

The larger study. 114 pre-service teachers (33.3% male sex; 77 Caucasian) from a large North American university were recruited for the project. Participants ranged in age from 18 to 41 years old (*M* = 23.1 years, *SD* = 4.86) and from 2.00 to 4.00 (out of 4.00) in self-reported GPA (*M* = 3.11, *SD* = 0.39, four missing values). 14 participants were sexual orientation or gender-identity minority participants (one non-binary participant; 13 non-heterosexual participants).

The samples were pre-service teachers taking an undergraduate education course. They were able to choose an experiment (out of two) to participate in or choose an alternative academic task if they opted out of study participation for course credits. When the participant chose to sign up for our study, they were given a demographics form to fill out prior to participating. When they completed the form and booked a timeslot for their session, they were assigned a participant number and assigned to one of two conditions: the experimental condition (learning task with the multimedia mobile app) and the control condition (gaming task). These conditions were further divided into two groups for counter-balancing the pre-test and post-test knowledge check questionnaires. In other words, each group from control and experimental conditions were given questions #1 to #14 on their pre-test knowledge-check questionnaire, and questions #14 to #28 on their post-test questionnaire; the other two groups were given #14 to #28 on the pre-test, and #1 to #14 on the post-test instead.

While the experimental group were given the EQH App (described below under the Multimedia Mobile App section), the control group was given a game called Plants vs. Zombies (PvZ). PvZ was an ideal control gaming task for several reasons. First, the game features content that is completely irrelevant to queer history and Edmonton. Second, the game is played with only a mouse, providing similar methods of interaction when compared to using the EQHA. Third, the game is a single player game and has minimal sources of distraction (e.g., does not have chat functions with other players), and hence provides a similar degree of non-social experience as the EQH App in a laboratory setting. Finally, PvZ has a low learning curve compared to more complicated games and can guarantee near identical and consistent experience between participants by letting all the participants start from the very first level in the Adventure mode.

When signing up, participants were given basic information about the study, which included that the study was about learning history through an app. It did not disclose any information on the specific content, or the specific topic of the app. Participants would therefore not have known the app was about the city they attend university in nor that it was about queer history. The consent form given just prior to beginning the experiment did mention the topic and the content of the app, regardless of what experiment group the participant was assigned to. The consent form can be viewed in appendix A.

The current study. The current study uses 20 participants' data collected from the larger study: ten participants with the highest proportional learning gains and ten participants with the lowest proportional learning gains. All of the participants were from the experimental group of the larger study. The 20 participants were 19 to 36 years old (M = 22.7, SD = 4.70), had GPAs between 2.70 and 3.50 (M = *3.10*, SD = 0.21, one missing value, 4.00 is the highest GPA obtainable). 13 of the 20 participants reported being female (sex), with one reporting to be non-binary in terms of gender identity. The seven male participants reported congruent biological/natal sex and gender identities. All of the participants reported being heterosexual. 14 of the 20 participants reported being Caucasian.

The samples were chosen based on proportional learning gains (also referred to as normalized learning gains). Compared to using 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). Marx and Cummings (2007) outline the formulae for calculating proportional learning gains, where *c* (normalized learning gains) equals to (post% – pre%) ÷ (100% - pre%), if the post-test score is higher than the pre-test score. If the post-test score is lower than the pre-test score, *c* would instead be equal to (post% - pre%) ÷ pre%. Finally, if the post-test and the pre-test scores are the same, the normalized learning gain is thought to be 0, and therefore dropped from the analysis.

Learners with a pre-test score of 9 and higher (maximum score of 14) were not eligible for the current analyses. If the learner scored high on the pre-test, it was assumed there was not much room for learning to occur. Hence, learners with a score of 60% or higher (9 out of 14 or higher) were thought to have scored beyond the threshold for this study, as this study was to focus on emotions and learning.

Multimedia Mobile App

The Edmonton Queer History App (EQH App) was developed by Harley, Liu, Ahn and colleagues (in press) using the Izi.TRAVEL content management system (CMS). Izi.TRAVEL allows users to create content that is suitable for exhibitions, tours, and general guides. The EQH

App was made through extensive interviews with local queer community members and research on queer history topics through media sources such as old newspapers or television recordings.

The EQH App features eight locations the learners can go through — the user may use a mobile device and go to these locations physically or opt to "move through" the locations virtually. Each location holds significant meaning towards queer history and culture in Edmonton and showcases the challenges the queer community endured, and the social changes that occurred throughout time. Each location features a multimedia experience, where users are presented with elements such as a video clip, an audio clip, historical photos, collages, or snapshots of artifacts with text captions for explanations, accompanied by multiple paragraphs explaining the location and the events that happened around it. Further, the user is given a digital map, and is able to enter "Street View" mode (offered by Google Maps) to look at how the location has changed in present times.

The app contains a tutorial video to teach users how to navigate within the app and has recommendations for optimal experience. The app also contains a glossary that explains important terms for understanding the app content. The app is accessible through any modern internet browser (e.g., Mozilla FireFox, Chrome, Safari) across multiple platforms (e.g., smart phones, laptops, etc.). The user interface (i.e., how the app looks) differs slightly across platforms but does not contain any significant changes, perhaps with the exception of how a user might hold the device, and how big the content would be when viewing. The learners in this study were given access to the app via Google Chrome on a desktop (left half of a 24" 1080p screen), and were instructed to use the mouse and keyboard to interact with the app. Figure 1 shows a mobile version of the app, while figure 2 shows a screenshot of what the learners would have viewed.

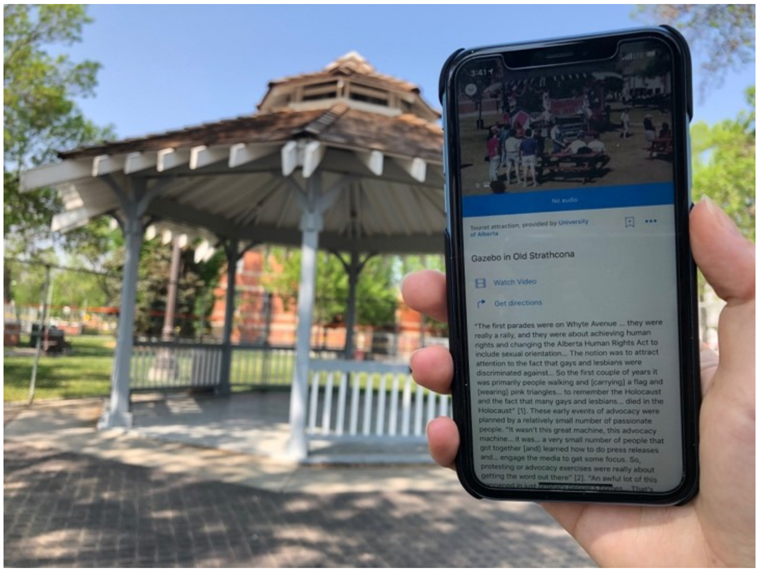


Figure 1. A photograph showing what a user might experience when using the mobile version of the app and visiting the locations physically.

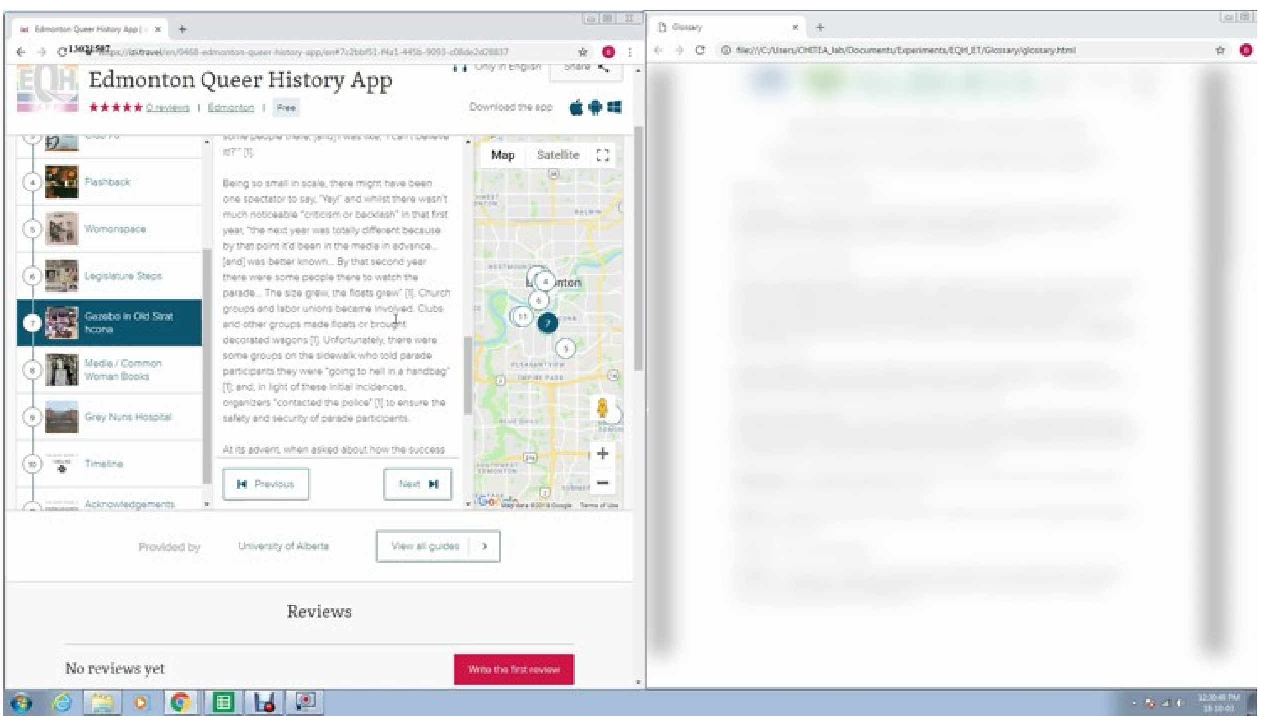


Figure 2. A screenshot showing the desktop browser version of the EQH App the learners of this study experienced. The left side of the screen shows the EQH App, while the right side of the screen shows a blurred glossary containing terms relevant to the EQH App. The inclusion of the glossary and the method of interaction (i.e., click on it to unblur the content) was part of larger study's design and is not the focus on this thesis.

Measures

The current study focused on a subset of the data channels used in the larger study. The larger study included measures such as electrodermal activity and heart rate through an Empatica E4 bracelet, and measurement of affective components such as emotions, empathy, and hope through various self-report questionnaires. While interesting questions can always be posed with various types and forms of data, I have selectively focused my attention on a few of the data channels to keep the scope of this thesis coherent and answer my research questions: Emotions from facial recognition software and video capture of learners' faces, attention through eye gaze behavior from an eye-tracker, and learnings gains from counter-balanced pre- and post-app interaction measures.

Attention. Eye gaze data were recorded through EyeLink 1000 and the accompanied software. Specifically, I was able to obtain a beta version of an EyeLink 1000 software directly from its developer, SR Research, that has screen-recording features. I was able to capture the computer monitor screen activity as a video that includes a blue dot indicating where the learner's gaze was. Figure 3 shows a screenshot for reference. While it is possible to pre-program EyeLink 1000 to detect gazes occurring within areas of interests (AoIs), this feature was not appropriate to apply to this experiment as the AoIs were not stagnant. That is, the AoIs for this experiment refer to areas defined around sentences which appear on the app. The app, however, presents the text in a box where the text will move around as the learner scrolls through; this makes it difficult to determine a pre-defined coordinate of the AoIs. Instead of relying on software to detect gaze duration within AoIs, this task was done through systematic coding through human labor. Details on the exact process are described below in the Procedures sub-section.

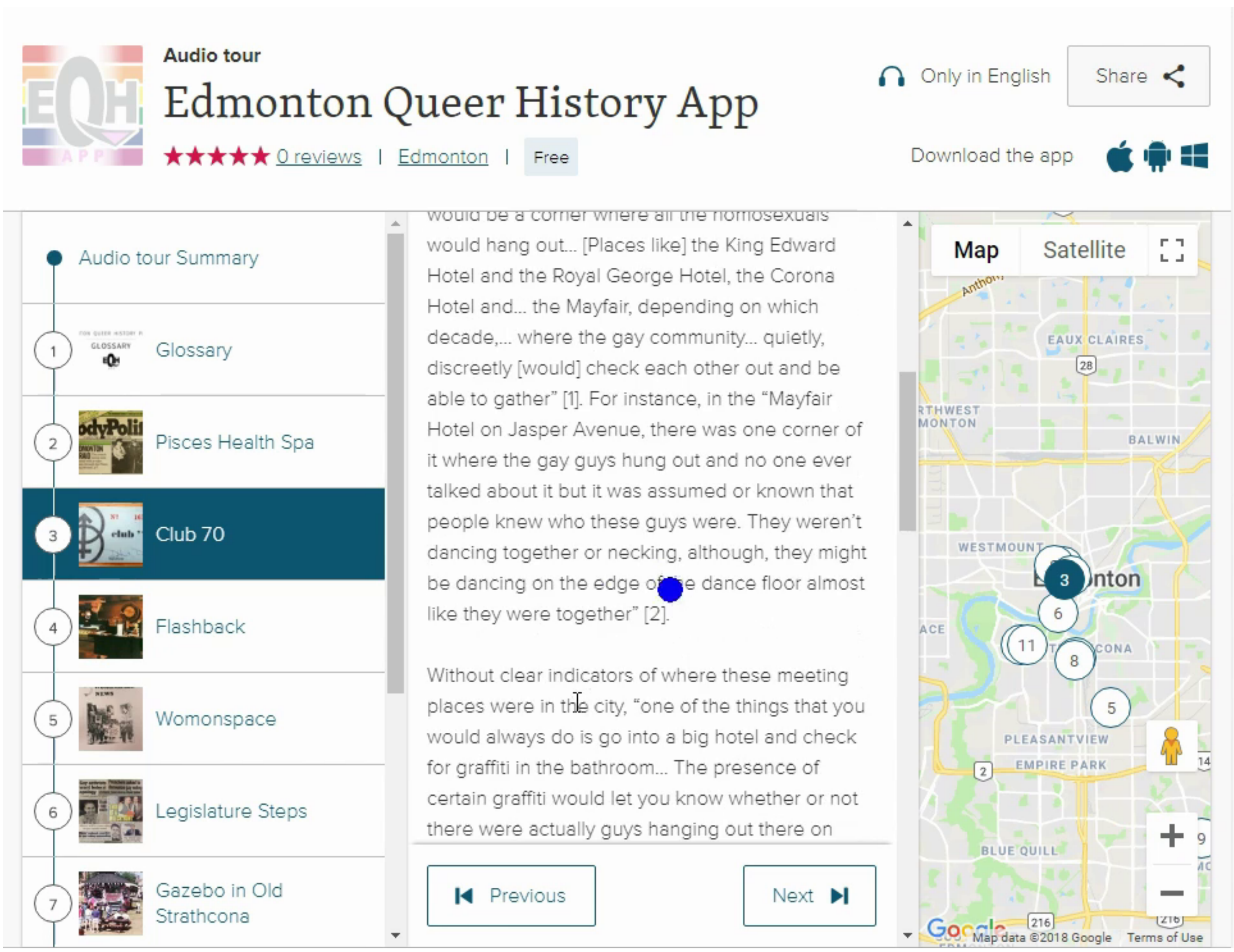


Figure 3. A screenshot of an eye-gaze captured video produced by EyeLink 1000. The blue dot indicates the participant's current gaze.

The EyeLink 1000 was set to Desktop Remote Mode and was set to use the 25mm lens to record the left-eye gaze monocularly. The experimenter conducted nine-point calibration/validation for each participant, and further conducted a drift check/correction. When the experimenter noticed that the eye-tracking was not optimal after the calibration, the experimenter paused the experiment and re-calibrated/validated the eye-tracker.

Emotions. Emotions were measured by referring to the output files of FaceReader 7. FaceReader is an automatic facial expression recognition software that uses machine learning to identify human expressions as discrete emotions. FaceReader 7 employs two processes for face classification: 1) face modeling and classification; and 2) deep face classification (Loijens, Krips, Grieco, van Kuilenburg, den Uyl, & Ivan, 2016). The former refers to a method based on Active Appearance Modeling and analyzes over 500 key points of the target face. Then, the face is fed into an artificial neural network that will classify the face into six basic emotions (or neutral), as described by Ekman (1992). Noldus reports that the artificial neural network was trained with over 10,000 images that were manually annotated by experts. The second method refers to a process where FaceReader does not conduct any face modeling, but rather directly recognizes expressions based on patterns from the image pixels. Noldus reports that this method can help analyze faces that are partly covered. The final output of a classification is the combination of the two methods' results.

Noldus 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; Liujens et al., 2016). They further reported that the addition of deep face classification in FaceReader 7 will make the accuracy even higher (Loijens et al., 2016). A study by Terzis, Moridis, and Economides (2010) examined FaceReader and their own researchers' agreement on categorizing emotions, and reported that disgust (70%) and anger (71%) showed relatively lower agreement ratings between the researchers and FaceReader, while neutral (99%) and happy (90%) showed the highest agreement rates. In addition, a more recent examination by Lewinski, den Uyl, and Butler (2014) report that FaceReader 6 had an 88% accuracy when identifying basic emotions based on prototypical facial expressions and was almost eligible for a Facial Action Coding System (FACS) certification with a score of 0.69 (0.70/1.00 needed minimum for FACS certification). When FaceReader 6 is compared to self-report measures, Harley and colleagues (2015) reported a high agreement (75.6%) between the two. It should be noted that this was the case when the emotions with similar valence and arousal levels (e.g., anger and frustration; fear and anxiety, etc.) were grouped together based on Pekrun's work (2011). In summary, while FaceReader is not perfect, like any measure of emotion, it is competitive when compared to human coding and the continuous improvement of the software is promising for accurately and automatically measuring discrete emotions.

For my thesis, each participant was calibrated through FaceReader for better accuracy. Specifically, another lab member trained in machine learning and FaceReader selected a section of the video that they subjectively perceived to be a neutral face. Then, the appropriate face model (i.e., Caucasian, Eastern Asian, African, South Asian, or Other) was chosen through subjective judgment. After the calibration, each participants' webcam videos were fed into FaceReader, where two types of log files were generated: the state log, and the detailed output log. The detailed log files have in-depth information such as arousal, accuracy, blinks, and so on, but I have opted to refer to the state log file, which contains timestamps of all the emotions detected by FaceReader. An emotion is considered to be detected when enough action units

(AUs – facial muscle groups that form expressions) are detected to be similarly active according to FaceReader's classification of discrete emotions. In summary, while the detailed log can provide more granular details on whether or not minute traces of certain facial expressions were present, the state log file contains all the necessary information to figure out when an emotion was actually expressed and how long the emotion lasted.

Learning gains. Proportional learning gains were measured by using a pre-test and post-test knowledge check questionnaire. All except for one question (which was a true of false question) were multiple choice questions with four choices. The questions were created by two major contributors to the development of the multimedia mobile app, specifically by the two members with expertise in LGBTQ+ rights and history. The internal reliability of the 28 questions was low with a Cronbach's Alpha score ($\alpha = 0.55$), as expected when targeting many different topics within the app (Rowe et al., 2017).

Due to counter-balancing, approximately half of the participants from the larger study were given test version "A", while the other half were given "B". When the participants were selected for this study, 12 of the participants took the version "A" of the knowledge check questionnaires, while the other eight participants took the version "B". Although the questions between the versions differed, they were intended to target similar topics and ideas, and have similar difficulty levels. Papenberg's (2018) minDiff R-package was used to run 500,000 iterations of an algorithm designed to minimize group differences in terms of total correct answers per questions (question difficulty as a numerical factor), and the historical locations the questions targeted (question topic as a categorical factor). The package was run based on a previous study that used these questions ($n = 57$; Harley, Liu, et al., in press). In terms of difficulty, test version "A" expects a mean score of 80.6% with a standard deviation of 17%

based on scores from a previous study. Likewise, test version “B” expects a mean score of 80.7% with a standard deviation of 17%, indicating near identical difficulty levels of the test items. The algorithm also assigned a similar number of questions targeting the same location. Here are a couple example questions seen in the questionnaire: What were the most popular kind of events Womonspace held? a) Book readings; b) Political rallies; c) Dinners and dances*; d) Home repair lessons; Which of the below was an outcome of the Pisces Bathhouse Raids? A) Increased activism*; B) Even greater efforts to conceal gay establishments; C) Decreased activism; D) The gay community feeling less vulnerable.

Procedures

Data collection. The experimental study took place in a laboratory setting, one participant at a time. Each participant was given a demographics form to complete prior to attending. Once at the experiment location, the experimenter escorted the participant into a small quiet room and obtained consent for the experiment. The experimenter then gave short instructions on how to interact properly with the measurement equipment (e.g., having proper posture for the eye-tracker), and briefed the learner, including whether they were interacting with a game or a history learning app. The participants were then given questionnaires. The control groups were given a game to play for 30 minutes. The experimental groups were 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). Finally, learners in both conditions were given the post-interaction questionnaires before they were debriefed and dismissed.

Data Analyses

Data cleaning and organizing. I used three forms of data in my analyses: 1) screen capture videos featuring learners' gaze; 2) FaceReader log files with facial expressions and timestamps; and 3) proportional learning gain scores.

Gaze Tracking Screen Capture Video. The 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, I came up with a systematic way of deciding whether or not a learner was looking at an AoI. First, I 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 methodology, specifically regarding the alignment of the gaze data with the facial expression data. If I were solely interested in finding out when a facial expression occurred, I could simply consult the log files FaceReader outputs. However, to figure out when a facial expression occurred in relation to gaze behavior, I needed to do the following: 1) consult the screen capture video, and figure out when a specific gaze instance occurs—I 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 me 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—I 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, I can now say it begins at 1:40:04 PM). 3) From the new time stamps obtained, I 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, I 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 from the process described above; the Windows 7 system clock by default only shows hours and minutes, and with additional tweaks I was able to make it show seconds, but not milliseconds. Because the smallest unit of time I can refer to are seconds, this poses a limitation regarding how granular I 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. Figure 4 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. Figure 5 is one 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.

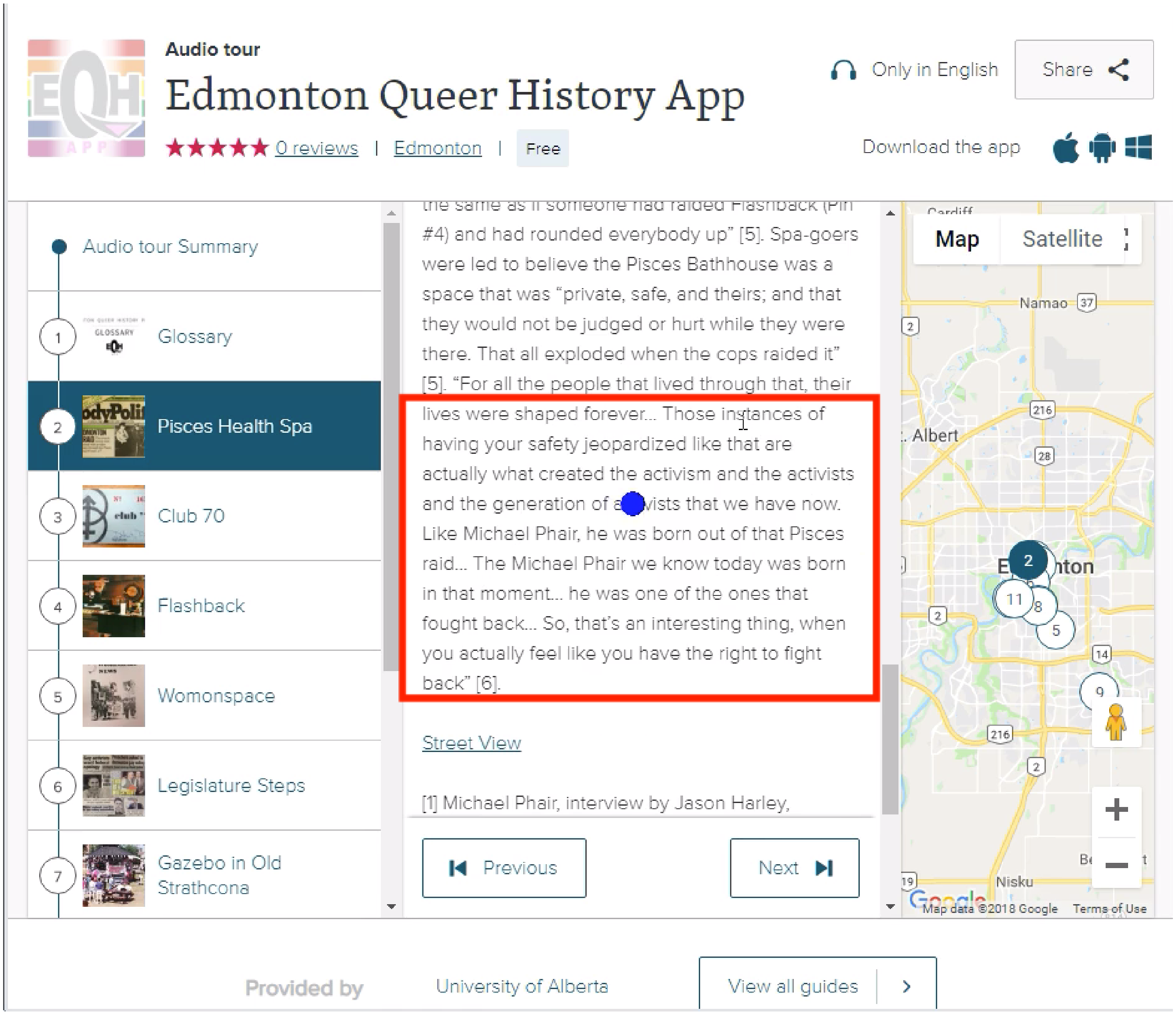


Figure 4. Screenshot that tells the coders where the AoI is. By referring to this screenshot and the gaze capture video with the added grids, it is possible to determine systematically whether or not the blue dot is inside the AoI.

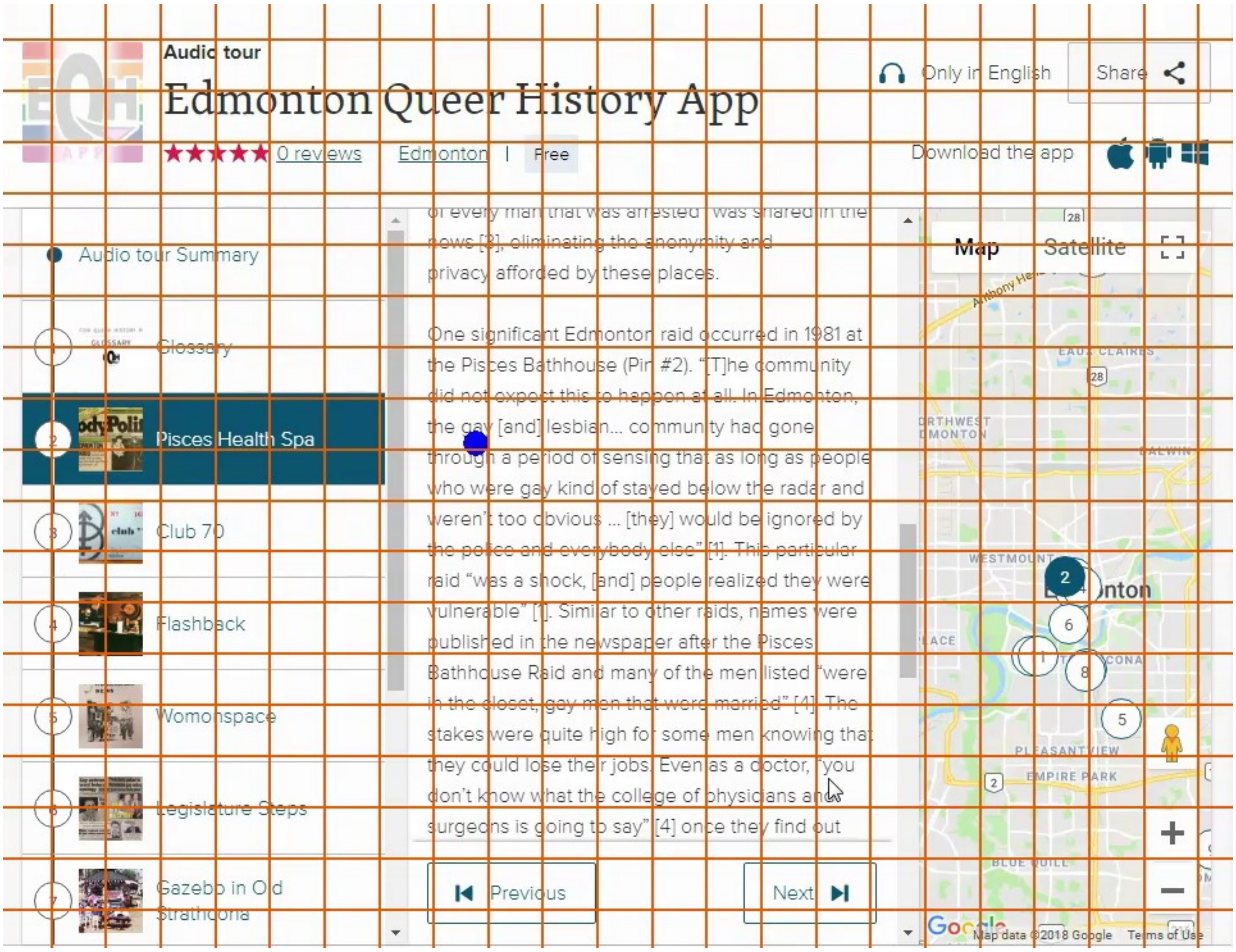


Figure 5. Screenshot of the EyeLink gaze screen capture video with grids added on. The added grid can help determine whether the gaze (blue dot) is inside the AoI or not.

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 framerate, 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 my direct supervision. The assistants first recorded each gaze instance from the screen capture videos. Then, they identified what timestamp had to be identified 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, 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 worst 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.

FaceReader log files. After identifying each instance of gaze towards the AoI, it was possible to align that information with the FaceReader state log file to identify whether or not an emotion was expressed while learners read critical parts of the app that had answers to the knowledge test questionnaire. I was able to create a spreadsheet that showed the durations of each emotion the learners expressed, and the dominant emotion that each learner expressed. The emotion "surprised" was dropped from analyses due to the vast majority of cases where FaceReader reported "surprised" occurring when a learner was yawning, or immediately after yawning. Further, "neutral" was not included in the analyses as the focus of this analyses was on the impact of emotions. Neutral facial expressions often vastly dominated recordings of the learning session (Harley et al., 2015).

In summary, by coding and organizing the raw data I had, I was able to create a spreadsheet that had each learner's data regarding: a) duration of learning; b) duration of total gaze durations towards AoIs; c) duration of each emotion expressed; d) duration of each emotion expressed while gazing at the AoIs; e) dominant emotions expressed during the learning session; f) dominant emotions expressed while gazing at AoIs; and g) proportional learning gains. The spreadsheet further contained demographic data. This data was used to conduct two phases of analyses: it was first used to investigate the frequency, duration and types of emotions during just the AoI gazes; it was then used to investigate the frequency, duration and types of emotions during the entire learning session. The second phase employed one-way ANOVA, Kruskal-Wallis H test, and the Mann-Whitney U test to investigate whether learners' dominant emotion profiles had an effect on learners' knowledge outcomes.

Chosen statistical tests. Statistical tests were chosen and the assumptions required for the tests were examined, as described below.

Dominant emotion profiles and learning outcomes. A one-way ANOVA was chosen to investigate whether there were group differences in proportional learning gains when learners were grouped by their dominant discrete emotions. The assumptions for a one-way ANOVA were first checked. There were no outliers in the data, as assessed by inspection of boxplots. Normalized learning gain score was normally distributed for anger, happy, and sad groups, as assessed by Shapiro-Wilk's test ($p > .05$). However, the scared group did not have a normal distribution ($p = .035$). While ANOVA tends to be robust towards marginal violation of the normality assumptions, Hae-Young (2013) reports that a Z skewness and Z kurtosis score between -1.96 and 1.96 provides enough confidence for such a test when dealing with small (less than 50) samples. Z skewness and kurtosis are calculated by dividing the original skewness/kurtosis by the standard error skewness/kurtosis respectively. However, despite the re-calculations, the Z kurtosis for the "scared" group (-2.282) was still not adequate for a one-way ANOVA. The Z skewness score was -0.004. Hence, a non-parametric alternative to one-way ANOVA, the Kruskal-Wallis H test was conducted. The Kruskal-Wallis H-test is a non-parametric counterpart of a one-way ANOVA and is appropriate to use when dealing with data with an asymmetric distribution (as is the case with this study; Chan & Walmsley, 1997; Hecke, 2012). Specifically, it does not require the assumption of normal distribution in the examined population (hence the sample) by using rankings.

Difference in learning gains depending on the dominant emotion categories. Discrete emotions (e.g., anger) can be categorized into a valence-activation category (e.g., negative-activating). By doing so, three categories can emerge from the data: Negative-activating (i.e., anger, scared); negative-deactivating (i.e., sad); and positive-activating (i.e., happy). A one-way ANOVA was chosen again to investigate whether there were differences in proportional learning gains when learners were grouped by their dominant emotions' valence-activation category. The assumptions for a one-way ANOVA were first checked. There were no outliers in the data, as assessed by inspection of boxplots. The normalized learning gain score was normally distributed for negative-deactivating and positive-activating groups, as assessed by Shapiro-Wilk's test ($p > .05$). However, the negative-activating group did not have a normal distribution ($p = .019$), and thus Kruskal-Wallis H test was chosen again. In addition to the H test, a one-way ANOVA was still run; the negative-activating group's Z skewness (-1.507) and Z kurtosis (-0.155) indicated an acceptable level of deviation from the normal distribution.

Results

This section presents the results to answer the proposed research questions. Analysis 1 focused on data related to AoIs, while Analysis 2 focused on data related to the entire learning session with the EQH App. It should be noted that I defined dominant emotion 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. The smallest difference between a dominant and non-dominant emotion observed was from PN10's 3.5 total seconds (0.14% duration of the entire session) of happiness and 1.6 total seconds (0.06% duration of the entire session) of sadness, with a difference of 1.9 seconds.

Analysis 1

Frequency, duration, and types of emotions during AoI gazes. Learners' overall gaze duration towards 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 emotions 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 emotion was 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 = 2,198 seconds, SD = 609). During those learning sessions, learners' total duration of expressing emotions ranged from 0 seconds to 155 seconds (M = 25.6 seconds, SD = 42.1). In other words, learners showed facial expressions for 0% to 11.1% of the time during gazing at the AoI (M = 1.35%, SD = 2.58). Similar to the first analysis, little emotion was detected during the learning session. However, in terms of the absolute frequency of emotions, more emotions were detected in the second analysis as expressions outside of the AoI gaze duration were also captured. While only 35% (7) of the learners had expressed any emotions during the AoI gazes in the first analysis, the second analysis showed that 90% (18) of the learners expressed emotions during the learning session. In terms of the dominant discrete emotions, anger accounted for 20% (4) of the learners, happy accounted for 15% (3) of the learners, sad accounted for 35% (7) of the learners, and scared accounted for 20% (4) of the learners. When these emotions are categorized accordingly to their valence and activation, 40% (8) of the learners had negative-activating dominant emotions (i.e., anger, scared), while 35% (7) of the learners had negative deactivating emotion (i.e., sad).

Finally, 15% (3) of the learners had dominant positive-activating emotion (i.e., happy). A summary of analysis 1 and 2 can be found in Table 1.

Table 1

Frequency of Dominant Emotion Profiles of Learners

Discrete Emotions / Valence-Activation Categories	Number of Learners: During AoI Gazes (% of total N)	Number of Learners: During Entire Session (% of total N)
Anger	2 (10%)	4 (20%)
Happy	1 (5%)	3 (15%)
Sad	2 (10%)	7 (35%)
Scared	2 (10%)	4 (20%)
Total Non-Neutral Emotions	7 (35%)	18 (90%)
Negative-Activating	4 (20%)	8 (40%)
Negative-Deactivating	2 (10%)	7 (35%)
Positive-Activating	1 (5%)	3 (15%)
Total Non-Neutral Category	7 (35%)	18 (90%)

Proportional learning gain scores and emotions.

Differences in learning gains depending on the dominant discrete emotions. I investigated whether or not there were statistically significant differences in learning gains depending on the expressed dominant emotions. Figure 6 shows a histogram of normalized learning gains, split by the dominant emotions expressed.

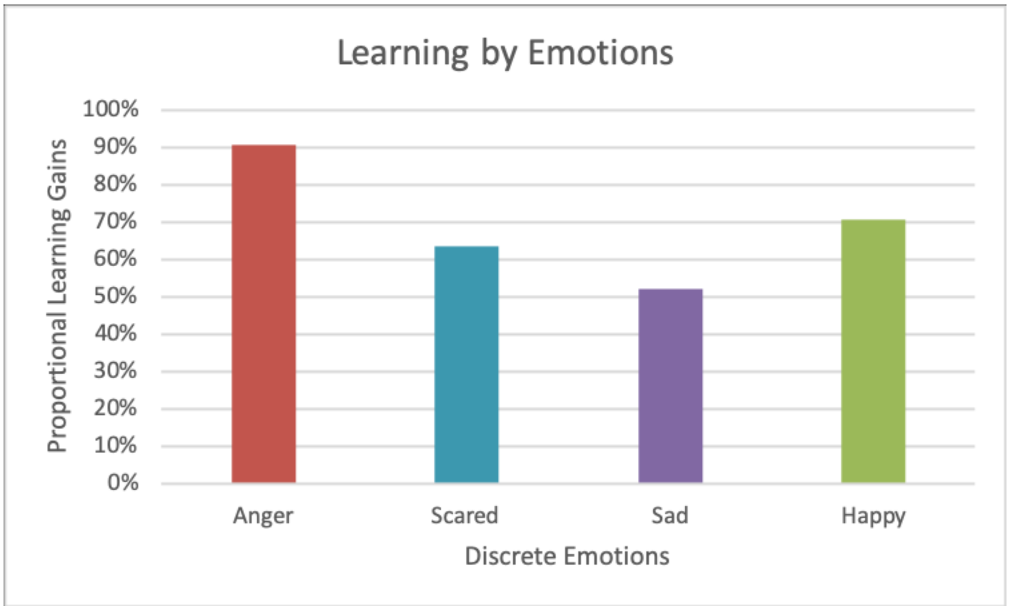


Figure 6. Histogram of normalized learning gains split by the dominant emotions expressed during the learning session.

The Kruskal-Wallis H test was conducted to determine if there were statistically significant differences between the “anger” (n = 4), “happy” (n = 3), “sad” (n = 7), and “scared” (n = 4) groups. The test did not allow inferences about differences in medians between the groups, due to the unsimilar shapes of the four groups’ distributions (inspected by boxplot displayed in Appendix B). Rather, the test allowed inferences about differences in distributions, lower/higher scores and/or mean ranks. The distributions of normalized learning gain scores were statistically different between groups, $\chi^2(3) = 8.33$, $p = .04$. Pairwise comparisons were performed using Dunn’s (1964) procedure with a Bonferroni correction for multiple comparisons. Statistical significance was accepted at the $p < .0083$ level for the post hoc analysis, and it revealed statistically significant differences in normalized learning scores between the sad group (mean rank = 6.43) and anger group (mean rank = 15.75, $p = .029$

[adjusted significance; see Table 3]). Table 2, 3 and figure 7 contains summarized results of the test.

Table 2

Summary of Kruskal-Wallis H Test for Proportional Learning Gains Split by Emotions

Discrete Emotions	Median Proportional Learning Gains (%)	N	Mean Rank
Anger	89.445	4	15.75
Happy	83.330	3	10.17
Sad	50.000	7	6.43
Scared	63.885	4	8.12
Total	81.665	18	

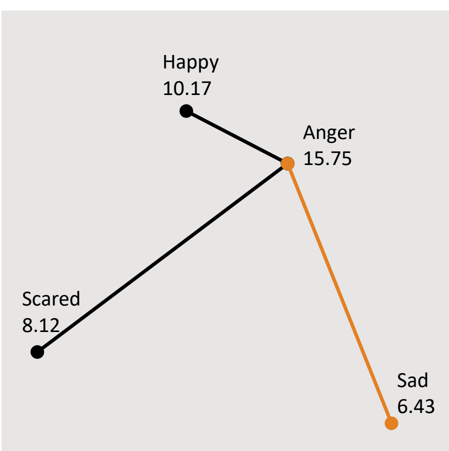


Figure 7. Graph of the pairwise comparisons. The yellow line indicates a significant difference while the black line shows non-significant differences.

Table 3

Summary of Pairwise Comparisons of the Emotion Groups

Sample Pairs	Test Statistics	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.
Sad-Scared	-1.696	3.303	-.514	.607	1.000
Sad-Happy	3.738	3.636	1.028	.304	1.000
Sad-Anger	9.321	3.303	2.822	.005	**.029**
Scared-Happy	2.042	4.024	.507	.612	1.000
Scared-Anger	7.625	3.726	2.046	.041	.244
Happy-Anger	5.583	4.024	1.387	.165	.992

Difference in learning gains depending on the dominant emotion categories. I further investigated whether or not there were statistically significant differences in learning gains depending on the valence-activation category of expressed dominant emotions. Figure 8 shows a histogram of normalized learning gains, split by the dominant emotion categories.

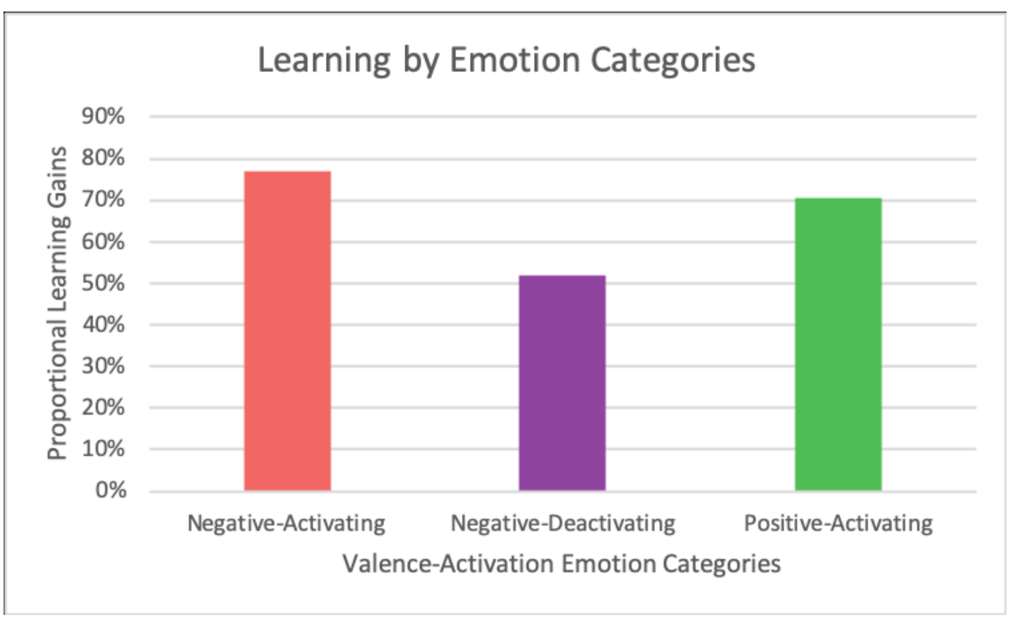


Figure 8. Histogram of normalized learning gains, split by the valence-arousal category of dominant emotions expressed during the learning session.

The Kruskal-Wallis H test was run to test for statistically significant differences in proportional learning gain scores between the "negative-activating" (n = 8), "negative-deactivating" (n = 7), and "positive-activating" (n = 3) groups. The test did not allow inferences about differences in medians between the groups, due to the unsimilar shapes of the three groups' distributions (refer to boxplot from Appendix C). Rather, the test allowed inferences about differences in distributions, lower/higher scores and/or mean ranks. However, the distributions of normalized learning gain scores were not statistically different between groups, $\chi^2(2) = 4.138$, $p = .126$.

Difference between negative-activating and negative-deactivating groups. The positive-activating group (i.e., learners with happy as their dominant emotion) had a very low sample size. Further analyses were therefore done to compare the normalized learning gain scores of just the negative-activating and negative-deactivating groups. A Mann-Whitney U test was run, as the negative activating-activating group's proportional learning gain score did not have a normal distribution as assessed by Shapiro-Wilk's test ($p = 0.019$). The median proportional learning gain score was not statistically significantly different between the two categories of negative emotions, $U = 11.500$, z = -1.934, $p = .053$.

Discussion

This thesis set out to examine what types of emotions were generated while learners engaged in learning queer history through a multimedia app, and whether learners' dominant emotions had a relationship with learning outcomes. The research questions were: RQ1a) How often did learners express an emotion while they read content directly related to the answers on the knowledge check questionnaire? b) What kinds of emotions did learners express when categorized by (i) discrete emotions and (ii) valence-activation dimensions? RQ2 Were there statistically significant group differences in proportional learning gains when learners were grouped by their dominant emotion (e.g., anger vs. happiness)? Further, these research questions were repeated for the entire learning sessions. I aimed to answer these questions by employing measurements backed by appropriate theoretical frameworks: eye-tracking to measure attention, automatic facial expression analysis to measure emotions, and a set of pre- and post-test questionnaire to measure proportional learning gains. The following sections will examine each of these research questions and discuss the results obtained in relation to the theories I have covered. I will then discuss the limitations of the current study. Finally, I will note the contributions this study makes and potential future directions to extends this program of research.

Research Question 1a: Analyses 1 and 2

My hypothesis that learners would have neutral expressions most of the time during learning was supported by the results. The lack of emotional facial expressions was especially the case with learners' frequencies of emotions during AoI gazes, where the majority of learners did not have a dominant emotion profile. Findings show that learners expressed emotions for roughly 1.5% of the time when reading important text content (AoIs) from the EQH App. Overall, the finding that neutral expressions were the most common while learning aligns with previous studies (Harley et al., 2015; Jarrell et al., 2016).

While little emotion was detected during AoIs, the vast majority of learners could be classified as having a dominant emotional profile over the course of the entire learning session.

This suggests that more emotionally-salient moments tended to be experienced outside of reading sentences directly related to the answers of the knowledge test. Overall, these descriptive findings reveal that the experience of learning queer history with the EQH App may have facilitated generally low-intensity emotions that did not necessarily lead to behavioral manifestations of emotion (i.e., through facial expressions). Indeed, while other studies have included manipulations to ensure specific emotions are elicited from learners, the AoIs in this study were not chosen to elicit a specific response from the learners, but instead to associate emotions with learning. That said, efforts were made to make the app emotionally engaging by, for example, providing rich historical contextualization from communication of a social frame of reference (Harley, Liu, et al., in press; van Drie & van Boxtel, 2008). Hence, while the contents within the AoIs did not necessarily trigger emotional responses, the overall content and the learning experience of EQH App fostered emotions from most learners over the course of their interaction with the app.

Theories and research can further explain the lack of emotions expressed in this study, or what would have perhaps led to more expressions. Previous work on the EQH App (Harley, Liu, et al., in press) revealed a high score of usability for the app — indicating that the learners had very little difficulty using it (e.g., ease of understanding of how to use the digital map, easily knowing what button clicks will lead to what). The CVT would posit that having little control over learning due to poor usability would have led to more frequent expression of emotions, most likely frustration or anger. The CVT would also posit that if the course credit awarded to study participations depended on the post-test score, learners may also have been more likely to express emotions on account of potentially different appraisals of task value.

Different types of achievement emotions may further explain my results. Pekrun and Stephens (2012) explained that achievement emotions include many different types of emotions, such as epistemic and social emotions. Research has explored such emotions including Muis and colleagues (2015) who looked at how potentially controversial information regarding climate change may impact learners' epistemic emotions. The EQH App presents LBGTQ+ related topics which may also be a sensitive or controversial topic to some. However, the multimedia app did not seem to elicit a great amount of behavioral expressions of emotions from participants — perhaps indicating that few learners found the topics controversial, or that the EQH App presented the material in a way that was not too provoking. The latter was an explicit design objective in the creation of the app. Further, Pekrun and Stephens' (2012) explanation of social emotions tending to be elicited in social situations may also contribute to explaining why learners had mostly neutral expressions: they were in an isolated environment with no social cues, and hence had no need to express emotions to other learners or instructors.

Lastly, it should be noted that unlike other studies that have utilized self-report measures, this study relied on measuring facial expressions to infer emotions. Automatic facial expression analysis programs, while having the advantage of being able to measure emotions in real-time, are less likely to capture subtle emotions compared to self-reports. This also implies that using repeated, concurrent self-report measures over the course of learners' session (Harley et al., 2015) with the EQH App may have revealed more emotions, as a previous study (Harley et al., in press) and preliminary analyses of retrospective self-report data from the current study (Harley, Ahn, & Liu, under review) suggests may exist.

Research Question 1b: Analyses 1 and 2

My hypothesis that learners would express more discrete emotions of happiness (positive-activating) than other emotions such as sadness (negative-deactivating) or anger (negative-activating) was not supported. Previous findings with the EQH App (Harley, Liu., in press; Harley, Poitras, et al., 2016), and similar apps for learning history in a Canadian city (Harley, Lajoie, et al., in press) have reported that learners were generally happy and enjoyed their learning experience with the apps. My results, however, show that based on the emotions observed over the course of the entire learning session, a sad dominant emotion profile was the most common (35%). When looking at valence-activation groups, leaners with negative-activating emotions (i.e., anger and scared) accounted for 40% of the learners. Positive emotions only accounted for 15% of the learners. It should be noted, however, that the *incidence* of emotions from facial expressions that informed dominant emotional profiles was low.

I believe that the negative emotions observed are 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 (scared) 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 and continue to face to this day may have been more emotionally impactful than the enjoyment of learning experienced in other studies. In sum, the participants may have felt such emotions on account of being emotionally invested and engaged with the content. If they did not care for social justice in their city, the progression of societal values in Canada, or learning with the EQH App, they may not have felt emotions such as anger. High appraisals of task value in a previous study and the current one with the EQH App support this interpretation (Harley, Liu, et al., in press; Harley et al., under review).

Research Question 2: Analyses 1 and 2

My hypothesis that learning performance would differ based on dominant emotion profiles was only partially supported. There were no statistically significant differences observed between learners with different dominant emotions, except for learners who expressed anger and sadness: learners with an angry dominant emotion profile had higher proportional learning gains compared to those with a sad dominant emotion profile. Further, when the learning gains were split by dominant emotion valence-activation categories, no statistically significant differences between the groups were found. When comparing just the negative-activating (i.e., anger, scared) and negative-deactivating (i.e., sad) groups, the result approached significance (p = .05), and had a large effect size (η_p^2 = .26). Key points from these findings are that not all negative emotions have the same association with learning; anger and sadness both have negative valences but led to opposite results. Relatedly, even emotions in the same valence-activation category (anger and scared) may differ in their association with learning: scared (anxiety) seemed to be more related with lower learning performance, while anger seemed to be more related to higher learning performance.

The CVT literature (Pekrun 2014), along with Loderer and colleagues' meta-analyses (in press) has shown that emotions such as anger have, on average, led to worse academic performance. My results contradicted this finding: participants who expressed anger as their dominant emotion statistically outperformed those who expressed sad and also outperformed participants with anxious and happy dominant emotions based on descriptive statistics). 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 my interpretation by stating that emotions such as anger can be beneficial to learning, provided that this emotion is resolved, and promotes engagement. I 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-activation emotions (happy) showed high learning performance, as expected by CVT.

The emotion "scared" (anxiety) is also negative and activating, and hence should have aided learning just like anger did, based on my arguments, so far. Pekrun (2014) specifically addresses the potential for this finding, as he 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 related to avoidance (Carver & Harmon-Jones, 2009). Anger is an emotion associated with one exerting themselves towards a situation. If a progress towards a goal is violated by something or someone, anger may be elicited and fuel motivation to directly intervene. Anxiety, however, is different in that it is an emotion that would promote avoidance away from a situation instead. A person who is angry about mistreatment would want to fight it;

a person anxious and scared towards mistreatment would want to flee from it. This difference may lead to the observed difference of the roles these two emotions played in learning. Learners who were angry may be more motivated to find out more and learn about the mistreatments and how the events in the EQH App unfolded. Learners who were anxious may be more motivated to skim and not overly invest themselves in finding out what happened exactly.

Limitations

No measurement is perfect, and the imperfect nature of the sophisticated tools I utilized in this study bring several limitations that should be acknowledged. This section also explains other limitations that are unrelated to the characteristics of the instruments that I used.

Eye-tracking. An eye-tracker was used to identify moments of relevance for further analyses with facial expression recognition software, while relying on the eye-mind hypothesis (Just & Carpenter, 1980). There are two limitations to this approach: first, the use of the eye-tracker itself is subject to error and noise. For example, the EyeLink 1000 came with the desktop mount (as opposed to other types of mounts that are compatible), which the manual states the optimal camera-eye distance is fixed at about 38cm. Due to this, the ideal eye-tracking scenario would be the participant placing their face in a plastic mount that would keep the distance and overall posture that is ideal for an eye-tracking session. This was not possible due to the equipment being uncomfortable to use for anything more than a short session (about 15 minutes). That is, participants' posture (despite reminders every now and then to sit up straight) shifted considerably during the experiment in some cases, introducing inaccuracies in the eye-gaze data that were captured. For example, the blue dot indicating gaze may have had less gaps between video frames if the data was closer to perfect.

Another example of noise in EyeLink data has to do with participants wearing glasses and makeup. Glasses can negatively impact the accuracy of eye-tracking, while makeup, especially mascara, has tended to make tracking much more difficult, even after successful calibration. Calibration and validation are still possible, but while a participant without glasses and makeup can afford to move around a little bit more, I found that the eye-tracker was especially unforgiving towards suboptimal eye-to-camera distance and angles when dealing with participants with glasses and eye makeup.

The second limitation to using eye-trackers was the limitations of the eye-mind hypothesis. For example, Anderson, Bothell, and Douglass (2004) reported evidence that gaze behavior did not reliably indicate retrieval processes, while Fox, Merwin, Marsh, McConkie and Kramer (1996) reported that pilot trainees in flight simulations suffered performance more when the peripheral vision was compromised as opposed to the actual focused gaze. While these studies do not invalidate the way I applied the eye-mind hypothesis, it does raise the question of whether gaze toward a learning material guarantees cognitive processes. Most who have experience in teaching students will likely agree that gaze alone is insufficient to guarantee learning-related cognitive processes; an observation that highlights the limitations of eye-tracking in this and other studies.

Facial expression recognition. Facial expression software was used to infer the participants' emotions. The expressions were captured through a webcam, and the interpretation of what facial expression meant in terms of emotions was done through FaceReader 7—a software that employs an artificial neural network to categorize facial expressions into emotions. FaceReader 7 was more tolerant towards participants' posture when analyzing the videos of their faces compared the EyeLink 1000. The main limitation regarding FaceReader 7 lies in the limitations of the theory of basic emotion (Ekman, 1992) used to classify emotions. Barrett (2017) explained that research on basic emotions used trained actors to demonstrate what the six prototypical facial expressions should look like based on Darwin's work. In other words, the six facial expressions that represent basic emotions were not created from real-world observation.

This limitation is handed down to measuring emotions in this study as FaceReader's artificial neural network is based on human coding of facial expressions. That is, the neural network attempts to perform as well as what the theory and methodology allows—if basic emotion theory designates a certain set of AU activations to be regarded as "happy", FaceReader can only work within that constraint and will not consider any other sets of AU activity to mean happy, or will always interpret that set of AU activation to be happy regardless of how the participant is actually feeling. In other words, FaceReader recognizing a facial expression to be categorized as happy, does not guarantee that person was indeed feeling happiness. For example, a person may smile which may indicate happiness when the person really feels anxious. Another good example of this limitation can be seen in this study where the vast majority of facial expressions labeled as "surprised" were in fact participants yawning. FaceReader's neural network identified a set of AU activation that matched what the basic theory of emotion would consider to be "surprised". In reality, the actual emotion was probably something far from surprise (e.g., bored) as the AU activity was from a yawn, as opposed to a genuine surprised expression.

Knowledge test questionnaires. The knowledge test questionnaires could be improved for future studies. Specifically, the list of questions used for the knowledge pre and post-tests contained six questions that had the answer choice of "All of the above" as the last choice; the correct answer for all of these six questions were "D) All of the above". Question design such as this may have led participants to guess quite well what the correct answers were. This may even explain the relatively high average pre-test scores (55%) of participants, with 37% (21) of them scoring higher than 60% on the pre-test. Further, the “all of the above” questions posed a significant challenge for defining AoIs. This is because solving these types of questions can be possible by reading multiple sections of the app and is even possible when not having read large chunks of the text. This is the opposite of many other questions in the questionnaire, where defining AoIs was relatively simple as they would belong to a specific section of a paragraph.

Other limitations. The generalizability of this study is limited by the size and nature of the sample. My 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 my 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.

It should also be noted that the small sample used in this thesis came from a non-random down-sampling procedure which is a limitation. Down-sampling was performed because of the labor-intensive nature of the AoI coding and an objective to examine differences in the emotions learners with high versus low proportional learning gains expressed. This line of analyses was abandoned, however, due to the low incidence of emotions overall, particularly for AoI gazes. Therefore, I merged the two groups of learners for the present analyses. Adding additional participants from the larger study, as described in the future directions, will help address these shortcomings.

This study has further limitations regarding ecological validity and external generalizability due to the samples and the laboratory-based learning scenario. The sample consisted of pre-service teachers who were taking a second-year university course. The primary motivation for participating in the study was likely for course credit. These characteristics call for consideration when considering the transferability of the findings to a different population of students or non-university population as well as to a differently incentivized or a non-incentivized sample. Even within the university population, the relatively short, non-social, low-stake nature of this multimedia learning task may make it so that the results differ in other learning scenarios.

Another limitation is the lack of incorporating more than one data channel for examining emotions. Specifically, measuring achievement emotions and emotions directed toward different object foci can further illuminate results. Instruments such as self-reports and interviews may be one way of obtaining such data (Harley, Liu, et al., in press).

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. There is value in examining an app that provides queer history learning opportunities, as there is a lack of attention toward this subject domain. For example, Alemdag and Cagiltay (2018) in their review of eye-tracking studies in multimedia learning research, reported that out of 58 studies, there were 46 studies related to hard-sciences (e.g., physics) or mathematics (e.g., geometry). There was only one study on history, which examined African history. Meanwhile, Queer history topics remain scarce in Canada, despite problems related to homophobia and transphobia persisting (StatsCanada, 2019). Queer history education can contribute to preventing homophobia and transphobia by introducing perspective taking necessary to dispel ignorance (Harley, Liu, et al., 2019). Research and learning materials related to queer history education is precious but scarce, and this study adds to that pool of resources.

Second, preliminary findings from this study suggest that typically undesirable emotions should not be blindly dismissed from instructional design, but rather more closely examined. 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 has provided preliminary evidence that perhaps 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.

Immediate future directions inspired by the limitation section include examining additional participants from the larger study to increase the sample size, eliminate potential influences from the non-random down-sampling, enhance generalizability, and potentially identify stronger and additional statistical relationships. Incorporating other data channels such as self-report measures represents another direction. Adding AoIs that contain elements designed to elicit strong emotional responses such as anger would extend this line of research as well by helping me further evaluate the explanations I advanced regarding the potential role of anger in social justice history education

Finally, another prominent future direction is the investigation of emotion regulation. Harley and colleague’s (2019) Integrated Model of Emotion Regulation in Achievement

Situations (ERAS) incorporated and extended Gross' (2015) process model of emotion regulation and Pekrun's (2006) CVT. By doing so, it can not only help explain why learning-related emotions may have been expressed, but also how these expressions were impacted by various regulatory processes. In other words, while emotions may be elicited, how much, and in what form they are expressed depends on emotion regulation processes. Exploring how learners employ different emotion regulation strategies (if they do at all) may help draw a clearer picture of what occurs in a learning scenario with the EQH App and similar apps. In other words, are university education students good at emotion regulation and hence show little emotion that is undesirable in learning from facial expressions? Do these students get better at emotion regulation as they go through their university career? At what point do these emotion regulation processes fail? If they had to study a subject matter they are disinterested in for a whole school year? If they were overwhelmed with other academic tasks, how would things change?

Conclusion

In this thesis, I investigated emotions expressed by learners using automatic facial recognition software from recordings of learners' sessions with a multimedia mobile queer history app. I also examined whether different dominant emotional profiles were associated with different learning outcomes. Preliminary results from this thesis indicated that learners tended to show little emotion, facially. An investigation of dominant emotion profiles revealed that learners expressed more negative-activating emotions (anger, anxiety), and negative-deactivating emotions (sadness) than positive-activating emotions (happiness). Learners whom expressed anger as their dominant emotion, had the highest learning performance; one that was statistically significantly different from learners with a sad dominant emotion profile. Findings also indicated that negative-activating emotions may impact performance differently based on the specific dominant emotion expressed: Learners who dominantly expressed anger had statistically significantly higher proportional learning gains than those who dominantly expressed anxiety. Preliminary findings also suggested that traditionally undesirable emotions in learning, such as anger, may aid learning, under the condition that these emotions are appeased and promote engagement towards learning. The study highlights the importance of investigating under-evaluated subject domains such as queer history and exploring whether anger may indicate higher levels of cognitive engagement and emotional investment in the learning activity, resulting in higher learning gains.

References

Alemdag, E., & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education*, *125*, 413–428. https://doi.org/10.1016/j.compedu.2018.06.023

Anderson, J. R., Bothell, D., & Douglass, S. (2004). Eye Movements Do Not Reflect Retrieval Processes: Limits of the Eye-Mind Hypothesis. *Psychological Science*, *15*(4), 225–231. https://doi.org/10.1111/j.0956-7976.2004.00656.x

Artino, A. R., Holmboe, E. S., & Durning, S. J. (2012a). Can achievement emotions be used to better understand motivation, learning, and performance in medical education? *Medical Teacher*, *34*(3), 240–244. https://doi.org/10.3109/0142159X.2012.643265

Artino, A. R., Holmboe, E. S., & Durning, S. J. (2012b). Control-value theory: Using achievement emotions to improve understanding of motivation, learning, and performance in medical education: AMEE Guide No. 64. *Medical Teacher*, *34*(3), e148–e160. https://doi.org/10.3109/0142159X.2012.651515

Barrett, L. (2017). *How emotions are made: the secret life of the brain*. Boston: Houghton Mifflin Harcourt.

Butz, N. T., Stupnisky, R. H., & Pekrun, R. (2015). Students' emotions for achievement and technology use in synchronous hybrid graduate programmes: A control-value approach. *Research in Learning Technology*, *23*. https://doi.org/10.3402/rlt.v23.26097

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

Clore, G. L., & Huntsinger, J. R. (2007). How emotions inform judgment and regulate thought. *Trends in Cognitive Sciences*, *11*(9), 393–399. https://doi.org/10.1016/j.tics.2007.08.005

Darwin, C., Cain, J. & Messenger, S. (2009). *The expression of the emotions in man and animals*. London: Penguin.

van Drie, J., & van Boxtel, C. (2008). Historical Reasoning: Towards a Framework for Analyzing Students' Reasoning about the Past. *Educational Psychology Review*, *20*(2), 87–110. https://doi.org/10.1007/s10648-007-9056-1

Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics, 6*, 241-252.

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.

Fox, J., Merwin, D., Marsh, R., McConkie, G., & Kramer, A. (1996). Information Extraction during Instrument Flight: An Evaluation of the Validity of the Eye-Mind Hypothesis. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *40*(2), 77–81. https://doi.org/10.1177/154193129604000215

Gross, J. J. (2010). The Future's So Bright, I Gotta Wear Shades. *Emotion Review*, *2*(3), 212–216. https://doi.org/10.1177/1754073910361982

Gross, J. J. (2015). The Extended Process Model of Emotion Regulation: Elaborations, Applications, and Future Directions. *Psychological Inquiry*, *26*(1), 130–137. https://doi.org/10.1080/1047840X.2015.989751

Gross, J. J., & Thompson, R. A. (2007). Emotion regulation: Conceptual foundations. In *Handbook of Emotion Regulation* (p. 22). Guilford Publications.

Harley, J. M., Ahn, B., & Liu, Y. (2019). *Can mobile apps be used to improve queer history education? An investigation of learning and emotions.* Manuscript submitted for publication.

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., Carter, C. K., Papaionnou, N., Bouchet, F., Landis, R. S., Azevedo, R., & Karabachian, L. (2016). 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*(2–3), 177–219. https://doi.org/10.1007/s11257-016-9169-7

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, B.T., Grace, A.P., & Lajoie, S.P. (2019). *A new source of hope: Learning about queer history.* Poster presented at the Canadian Psychological Association (CPA), Halifax, Canada.

Harley, J. M., Liu, Y., Ahn, T.B., Lajoie, S. P., Grace, A., Haldane, C., Whittaker, A., McLaughlin, B. (in press). I've Got This: Fostering Topic and Technology-related Emotional Engagement and Queer History Knowledge with a Mobile App. *Contemporary Educational Psychology*, 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. https://doi.org/10.1080/00461520.2019.1587297

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

Harp, S. E., & Mayer, R. E. (1998). *How Seductive Details Do Their Damage: A Theory of Cognitive Interest in Science Learning*. 21.

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

Izard, C. E. (2007). Basic Emotions, Natural Kinds, Emotion Schemas, and a New Paradigm. *Perspectives on Psychological Science*, *2*(3), 260–280. https://doi.org/10.1111/j.1745-6916.2007.00044.x

Jaques, N., Conati, C., Harley, J. M., & Azevedo, R. (2014). Predicting Affect from Gaze Data during Interaction with an Intelligent Tutoring System. In S. Trausan-Matu, K. E. Boyer, M. Crosby, & K. Panourgia (Eds.), *Intelligent Tutoring Systems* (Vol. 8474, pp. 29–38). https://doi.org/10.1007/978-3-319-07221-0_4

Johnson, W. L., & Lester, J. C. (2016). Face-to-Face Interaction with Pedagogical Agents, Twenty Years Later. *International Journal of Artificial Intelligence in Education*, *26*(1), 25–36. https://doi.org/10.1007/s40593-015-0065-9

Just, M. A., & Carpenter, P. A. (1980). *A Theory of Reading: From Eye Fixations to Comprehension*. 26.

Kaplan, H., & Sadock, B. (1991). *Synopsis of Psychiatry* (6th ed.). Baltimore: Williams & Wilkins.

King, R. B., & Gaerlan, M. J. M. (2014). High self-control predicts more positive emotions, better engagement, and higher achievement in school. *European Journal of Psychology of Education*, *29*(1), 81–100. https://doi.org/10.1007/s10212-013-0188-z

Knörzer, L., Brünken, R., & Park, B. (2016). Facilitators or suppressors: Effects of experimentally induced emotions on multimedia learning. *Learning and Instruction*, *44*, 97–107. https://doi.org/10.1016/j.learninstruc.2016.04.002

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

Liew, T. W., Tan, S.-M., & Jayothisa, C. (2013). *The Effects of Peer-Like and Expert-Like Pedagogical Agents on Learners' Agent Perceptions, Task-Related Attitudes, and Learning Achievement*. 13.

Linnenbrink-Garcia, L., & Pekrun, R. (2011). Students' emotions and academic engagement: Introduction to the special issue. *Contemporary Educational Psychology*, *36*(1), 1–3. https://doi.org/10.1016/j.cedpsych.2010.11.004

Loderer, K., Pekrun, R., & Lester, J. C. (in press). 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.

MacLean, P. D. (1990). *The Triune Brain in Evolution: Role in paleocerebral functions*. New York: Plenum.

Marx, J. D., & Cummings, K. (2007). Normalized change. *American Journal of Physics*, *75*(1), 87–91. https://doi.org/10.1119/1.2372468

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

Moreno, R., & Mayer, R. (2007). Interactive Multimodal Learning Environments: Special Issue on Interactive Learning Environments: Contemporary Issues and Trends. *Educational Psychology Review*, *19*(3), 309–326. https://doi.org/10.1007/s10648-007-9047-2

Noteborn, G., Bohle Carbonell, K., Dailey-Hebert, A., & Gijselaers, W. (2012). The role of emotions and task significance in Virtual Education. *The Internet and Higher Education*, *15*(3), 176–183. https://doi.org/10.1016/j.iheduc.2012.03.002

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., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). *The Control-Value Theory of Achievement Emotions: An Integrative Approach to Emotions in Education*. 24.

Pekrun, R., Goetz, T., Daniels, L. M., Stupnisky, R. H., & Perry, R. P. (2010). Boredom in achievement settings: Exploring control–value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology*, *102*(3), 531–549. https://doi.org/10.1037/a0019243

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002a). Academic Emotions in Students' Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research. *Educational Psychologist*, *37*(2), 91–105. https://doi.org/10.1207/S15326985EP3702_4

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002b). Positive Emotions in Education. In *Positive Emotions in Education* (Vol. 1). https://doi.org/10.1093/med:psych/9780198508144.003.0008

Pekrun, R., Hall, N. C., Goetz, T., & Perry, R. P. (2014). Boredom and academic achievement: Testing a model of reciprocal causation. *Journal of Educational Psychology*, *106*(3), 696–710. https://doi.org/10.1037/a0036006

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

Pekrun, R., & Stephens, E. J. (2012). Academic emotions. In K. R. Harris, S. Graham, T. Urdan, S. Graham, J. M. Royer, & M. Zeidner (Eds.), *APA educational psychology handbook, Vol 2: Individual differences and cultural and contextual factors.* (pp. 3–31). https://doi.org/10.1037/13274-001

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

Reschly, A. L., Huebner, E. S., Appleton, J. J., & Antaramian, S. (2008). Engagement as flourishing: The contribution of positive emotions and coping to adolescents' engagement at school and with learning. *Psychology in the Schools*, *45*(5), 419–431. https://doi.org/10.1002/pits.20306

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, *6*(2), 135–147. https://doi.org/10.1016/j.edurev.2010.12.001

Rowe, E., Asbell-Clarke, J., Baker, R. S., Eagle, M., Hicks, A. G., Barnes, T. M., Brown, R. A., 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

Scherer, K. R. (2005). What are emotions? And how can they be measured? *Social Science Information*, *44*(4), 695–729. https://doi.org/10.1177/0539018405058216

Schutz, P. A., & Pekrun, R. (2007). Introduction to Emotion in Education. In *Emotion in Education* (pp. 3–10). https://doi.org/10.1016/B978-012372545-5/50002-2

Shuman, V., & Scherer, K. R. (2014). Concepts and Structures of Emotions. In *International Handbook of Emotions in Education*. https://doi.org/10.4324/9780203148211.ch2

Smilek, D., Carriere, J. S. A., & Cheyne, J. A. (2010). Out of mind, out of sight: Eye blinking as indicator and embodiment of mind wandering. *Psychological Science, 21*(6), 786-789.

Spering, M., Wagener, D., & Funke, J. (2005). Brief report: The role of emotions in complex problem-solving. *Cognition & Emotion*, *19*(8), 1252–1261. https://doi.org/10.1080/02699930500304886

Stark, L., Malkmus, E., Stark, R., Brünken, R., & Park, B. (2018). Learning-related emotions in multimedia learning: An application of control-value theory. *Learning and Instruction*, *58*, 42–52. https://doi.org/10.1016/j.learninstruc.2018.05.003

Stats Canada (2019). Police-reported hate crime in Canada. Retrieved from: https://www.statcan.gc.ca/pub

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

Terzis, V., Moridis, C. N., & Economides, A. A. (2010). Measuring Instant Emotions During a Self-assessment Test: The Use of FaceReader. *Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research*, 18:1–18:4. https://doi.org/10.1145/1931344.1931362

Um, E. "Rachel," Plass, J. L., Hayward, E. O., & Homer, B. D. (2012). Emotional design in multimedia learning. *Journal of Educational Psychology*, *104*(2), 485–498. https://doi.org/10.1037/a0026609

Wang, H., Chignell, M., & Ishizuka, M. (2006). Empathic Tutoring Software Agents Using Real-time Eye Tracking. *Proceedings of the 2006 Symposium on Eye Tracking Research &Amp; Applications*, 73–78. https://doi.org/10.1145/1117309.1117346

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6).

Wojciechowski, R., & Cellary, W. (2013). Evaluation of learners' attitude toward learning in ARIES augmented reality environments. *Computers & Education*, *68*, 570–585. https://doi.org/10.1016/j.compedu.2013.02.014

Appendices

Appendix A

Consent Form Used for the Study.

Jason M. Harley, Ph. D.
Assistant Professor, Department of Educational Psychology
5-139 Education North
University of Alberta
Edmonton, AB T6G 2G5
Tel: (780) 492–9170 E-mail: jharley1@ualberta.ca

Fostering Historical Reasoning, Hope, Empathy, Emotional Engagement and Queer History Awareness with a Mobile App

Attention: ____
[Research Participant's Name]

You are invited to participate in this study for which support has been obtained through a Social Sciences & Humanities Research Council of Canada Insight Development Grant. The objective of this research is to evaluate a new app that was developed in the first phase of this research grant. Participants in this study are assigned to one of two conditions which determines whether you will interact with the Edmonton Queer History App during your time in the lab or whether information about how to download and use it for free, at your convenience, will be shared with you before you leave the lab. Participants who do not interact with the app during the lab-based session will play a game, instead.

The Edmonton Queer History App is designed to teach LGBTQ+ and non-LGBTQ+ users about locations in the city of Edmonton that represent important historical locations for queer history, including the development of LGBTQ+ rights and freedoms. Our evaluation of the app includes collecting data related to your knowledge of queer history, your emotional engagement with the app, and empathy toward LGBTQ+ individuals.

Your participation involves either the use of the EQH app on a desktop computer in order to learn about Queer history locations in Edmonton or playing a game. We will gather information pertaining to what you know, and are thinking and feeling through the use of surveys administered before and after the session. You will interact with the experimenter who will help you to navigate the app/game and answer procedural questions. These interactions will be recorded. Specifically, we will capture how you interact with the app/game by using screen-recording software to record your interactions with the app/game. We will also use a video camera to record a video of your face to analyse your facial expressions to better understand your emotional responses to the app/game. An electrodermal activation device (bracelet) will be used measure your emotional arousal. An eyetracker will be used to capture your gaze behavior. After you have completed the study you will have a more complete explanation of the methods and purpose during the debriefing session.

Information (i.e., data) gathered from this study will be used to help validate and revise, if necessary, the app. It is anticipated that your participation in this study will take up to two hours.

.

Jason M. Harley, Ph. D.
Assistant Professor, Department of Educational Psychology
5-139 Education North
University of Alberta
Edmonton, AB T6G 2G5
Tel: (780) 492–9170 E-mail: jharley1@ualberta.ca

As a research study participating in the Department of Educational Psychology's Participant Subject Pool, you will be compensated for your participation in this two-hour research study with 5% course credit toward the final grade of the associated course.

You are invited to sign this consent letter in the space provided below once you read the following guidelines for participation:

- As a research participant, you are asked to provide signed informed consent in order to take part in this research.
- You have the right to refrain from answering any particular questions that make you uncomfortable.
- Your participation in this study is voluntary. You will have the right to opt out of the research at any time without penalty, and data can be withdrawn from the study upon your request at any time during your participation in the study and/or up to one week following your participation in the study.
- Processes to provide accuracy of data, security, confidentiality, and anonymity are implemented in the design of the study. Security and confidentiality measures will be implemented, including the back up of data, secure storage of tapes, and a plan for deleting electronic and taped data.
- Only Dr. Harley as the principal investigator (University of Alberta) and his collaborators, Dr. Andre Grace (University of Alberta), Dr. Kyle Mathewson (University of Alberta), Dr. Susanne Lajoie (McGill University), Dr. François Bouchet (Sorbonne Université, Laboratoire d'Informatique de Paris 6) and Dr. Eric Poitras (University of Utah), and their postdoctoral fellows, research assistants, and transcribers, all of whom are required to sign confidentiality agreements, will have access to data and information.
- You agree that my team and I can use information in secondary writing beyond the research report, which includes such writing as conference papers, book chapters, or journal articles. The same ethical considerations and safeguards will apply to secondary uses of data.
- Upon your request, you will be provided with a copy of the research report culminating from this study.
- We may use the data we get from this study in future research, but if we do this it will have to be approved by a Research Ethics Board.

If you have questions, please email Dr. Yang Liu at chitealab@ualberta.ca (supervised by Dr. Jason M. Harley).

CHI TEA
LABORATORY

.

Jason M. Harley, Ph. D.
Assistant Professor, Department of Educational Psychology
5-139 Education North
University of Alberta
Edmonton, AB T6G 2G5
Tel: (780) 492–9170 E-mail: jharley1@ualberta.ca

The plan for this study has been reviewed for its adherence to ethical guidelines and approved by Research Ethics Board 1 at the University of Alberta (Pro00084257). For questions regarding participant rights and ethical conduct of research, contact the Research Ethics Office at (780) 492-2615.

I have read this form and the research study has been explained to me. I have been given the opportunity to ask questions and my questions have been answered. If I have additional questions, I have been told whom to contact. I agree to participate in the research study described above and I will receive a copy of this consent form after I sign it. I provide my own independent consent to participate in this research:

Participant's Name (Print):____________________________________
Date: ____________________________________

Signature: ____________________________________

Additional (optional) consent. Please read the agreements below and circle "Yes" or "No" to indicate acceptance or not, to each of these items:

Yes/No I consent to have my audio-records that will be gathered during this study used (e.g., played) during dissemination of the results of this study at conferences and other venues (e.g., educational and training sessions)

Yes/No I consent to have my video-records (face video and screen recording) that will be gathered during this study used (e.g., played or shown) during dissemination of the results of this study at conferences & other venues (e.g., academic journals, educational & training sessions)

Yes/No I consent to have my electrodermal activation-records of my emotional arousal that will be gathered during this study used (e.g., played or shown) during dissemination of the results of this study at conferences & other venues (academic journals, educational & training sessions)

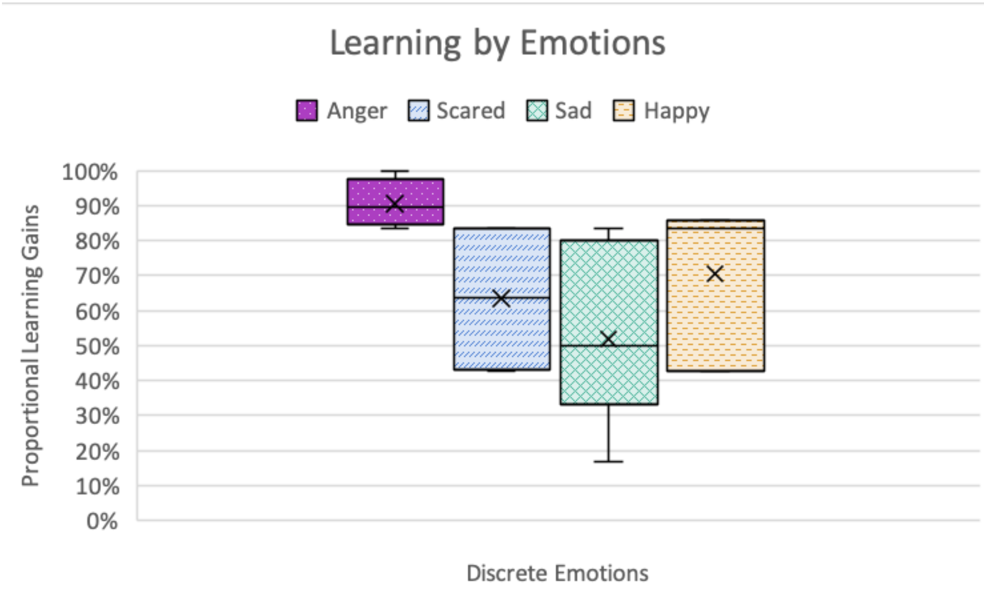
Yes/No I consent to have my eye-tracking records (gaze behavior) that will be gathered during this study used (e.g., played or shown) during dissemination of the results of this study at conferences and other venues (academic journals, educational and training sessions)

Participant's Name (Print):	Researcher's Name (Print):
Date:	Date:
Signature:	Signature:

CHI TEA LABORATORY

Appendix B

Boxplot of normalized learning gains, split by the dominant emotions expressed during the learning session.



Appendix C

Histogram of normalized learning gains, split by the valence-arousal category of dominant emotions expressed during the learning session.