Title	中国高等美術教育における時間ベース·メディアアートの学生鑑賞に対する感情的再評価の影響
Author(s)	馮, 浩城
Citation	
Issue Date	2025-12
Type	Thesis or Dissertation
Text version	ETD
URL	https://hdl.handle.net/10119/20318
Rights	
Description	Supervisor: 由井薗 隆也, 先端科学技術研究科, 博士

Doctoral Dissertation

THE INFLUENCE OF EMOTIONAL REAPPRAISAL ON STUDENTS' APPRECIATION OF TIME-BASED MEDIA ART IN CHINESE HIGHER ART EDUCATION

FENG Haocheng

Supervisor YUIZONO Takaya

Graduate School of Advanced Science and Technology
Japan Advanced Institute of Science and Technology
(Knowledge Science)

December 2025

Abstract

This dissertation investigates how emotional reappraisal during Time-Based Media Art (TBMA) appreciation influences students' emotional engagement, artistic understanding, interest, and liking in Chinese higher art education. It examines the dynamic processes through which repeated viewing and reflection influence emotional transformation and aesthetic appreciation. Emotion recognition technology was employed as a methodological instrument to capture and analyze these affective changes, thereby supporting a broader examination of how structured emotional reappraisal contributes to deeper and more reflective learning experiences in TBMA contexts.

This dissertation consists of three interrelated studies that collectively explore how emotional reappraisal during TBMA appreciation influences students' emotional and cognitive engagement. Study 1 investigated students' perceptions of the ethical and practical feasibility of integrating emotion recognition into TBMA education. The findings revealed conditional acceptance depending on factors such as device use, duration, transparency, and voluntary participation. These results provided a concrete foundation for the experimental design of the subsequent studies. Study 2 adopted a mixed-method approach that combined facial expression recognition with self-reported textual analysis to examine emotional dynamics during repeated viewing of TBMA works. By comparing students' immediate emotional responses with those observed after a temporal interval, the study traced how emotional responses developed across repeated viewing. The results showed that dominant emotions remained relatively stable over time, whereas non-dominant emotions became more diverse and more deeply elaborated, suggesting that guided repeated viewing supported processes of emotional reappraisal alongside cognitive regulation. Study 3, building on the preceding findings, further examined the relationship between emotional patterns and students' understanding, interest, and liking. Linear mixed model analysis revealed that, under certain conditions, greater emotional richness and intensity were associated with higher levels of understanding, interest, and liking. In addition, difference-score regression analyses indicated that increases in non-dominant facial emotions and the density of emotional words in self-reported texts positively predicted improvements in understanding, though not in interest or liking. These three studies form a coherent research progression. This progression begins with clarifying ethical feasibility in Study 1, moves to analyzing emotional processes in Study 2, and concludes

with examining their educational implications in Study 3.

These three studies reposition emotional reappraisal as a temporal mechanism of art appreciation, demonstrating that art learning is not a purely rational process but an affective and cognitive cycle involving reflection. Methodologically, this dissertation develops an empirically grounded measurement approach that combines facial expression analysis and text-based emotion analysis, together with a data handling procedure designed to respect contextual and ethical constraints, thereby enabling transparent, interpretable, and reproducible emotion-related evidence in TBMA education. From a course design perspective, guided re-viewing is proposed as a pedagogical strategy, employing temporally spaced re-viewing and guided reflection tasks to intentionally increase emotional diversity and depth, with the aim of enhancing understanding rather than merely stimulating preference.

Overall, this dissertation advances research across Knowledge Science, art education, and Time-Based Media Art by articulating emotional reappraisal as a guided reflective process supported by temporal spacing, which connects affective experience with cognitive development during art appreciation. Emotion is thus reframed not as a peripheral or purely experiential aspect of aesthetic engagement, but as a constitutive component in the formation of meaning and interpretation.

Keywords: time-based media art, emotion recognition, artistic aesthetics, guided re-viewing, higher art education.

Acknowledgment

I would like to express my heartfelt gratitude to all those who have contributed to the successful completion of this dissertation. The journey toward earning a doctoral degree is not only a testament to individual effort but also a reflection of the support and guidance provided by many.

First and foremost, I am deeply grateful to my esteemed supervisor, Professor Yuizono, whose expertise, insight, and patience have had a lasting impact on my doctoral experience. Thank you for giving me the opportunity to pursue my Ph.D. and for your continuous support through our weekly meetings, where you carefully followed my progress and helped me overcome every difficulty. Whether in academic research or in daily life, I have always felt your kindness and encouragement. Your dedication to mentoring and your thoughtful feedback have guided me not only in developing my research but also in improving the way I think and work as a researcher. I am also sincerely thankful to Wang Tzu-Yang sensei, my research advisor, who provided numerous insightful and practical suggestions that greatly strengthened my research methodology. His constructive comments and generous guidance helped me refine my approach and maintain clarity and rigor throughout the study. Throughout my doctoral process, the wisdom, patience, and support from both Professor Yuizono and Wang Tzu-Yang sensei have been invaluable.

I would also like to extend my sincere thanks to Professor Miyata for his thoughtful suggestions on my minor research project, and to my second supervisor, Professor Kim, for her invaluable guidance. My heartfelt appreciation goes to the members of my dissertation committee, Professor Nishimura, Professor Fujinami, and Professor Kanai, for their constructive comments and support during the entire process. Special thanks are due to the faculty and friends of Dalian Polytechnic University and Shenyang Normal University, whose support was instrumental in conducting the experimental component of this dissertation.

Reflecting on this doctoral journey, I am deeply aware of the personal and academic transformation it has brought. The process has taught me not only how to conduct rigorous research, but also how to think critically and persevere through uncertainty and doubt. There were moments of challenge and frustration, but they were matched by growth, resilience, and a stronger sense of purpose.

Finally, I dedicate this dissertation to my father and mother. Their unwavering love, guidance, and the values they taught me have always influenced my path. For many years I have lived overseas and could not stay by their side, meeting only through video calls. For that I feel truly sorry, and I am deeply grateful for their understanding and support. To my parents, I say: your son has made it.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

1.1 Research Background

1.1.1 Time-Based Media Art in Higher Education

The allocation of instructional time and the effectiveness of pedagogical delivery have long been central concerns in research on arts education (Scheerens et al., 2013).

In higher education, time has traditionally been conceptualized through structured divisions, such as course duration, weekly schedules, and semester or academic year planning. However, within contemporary curriculum design, time should not be regarded solely as a rigid framework; rather, it should function as an active agent in influencing the rhythm, sequencing, and depth of student engagement (Mittermeier and Benade, 2024; Rappleye and Komatsu, 2016). The allocation of time in educational programs plays a critical role in influencing students' cognitive and affective responses, particularly in disciplines that rely heavily on experiential learning, such as art education (Leek et al., 2024). Current theoretical frameworks in art appreciation emphasize temporality, audience engagement, and enduring aesthetics (Leder, 2013; Leder and Nadal, 2014). As a result, recent years have seen a notable expansion of research and exhibition related to time-based media art (TBMA) within higher education. This trend reflects broader transformations in media technologies and interdisciplinary pedagogical approaches (Perignat and Katz-Buonincontro, 2019). A growing number of universities and arts institutions have developed specialized programs. Leading institutions such as the Rhode Island School of Design (RISD) and the MIT Media Lab have integrated TBMA into disciplines including performance studies, media art, and experimental film (Sun et al., 2023). Likewise, higher education institutions in China and Japan have initiated programs and curricular discussions related to TBMA, with sustained efforts to align new media art education with the demands of the digital age (Spielmann, 2012; Zhao, 2025).

However, there remains a limited body of empirical research investigating how structured instructional time affects students' emotional interpretations of art. In parallel, as TBMA gains increasing visibility within academic art education, several persistent challenges have emerged. One of the most pressing concerns is the ephemerality of time-based works, which complicates both assessment and documentation (van de Vall, 2022). Unlike traditional artworks, TBMA projects often take the form of performance or digital media, requiring alternative evaluative frameworks such as audience engagement metrics or rapid-response interaction analysis (Nikolai et al., 2019; Shao and Wu, 2024). Another long-standing issue involves infrastructural and resource limitations associated with TBMA education. The field demands high-performance computing facilities (Yantaç et al., 2011), motion-capture laboratories, and specialized projection equipment, which present financial burdens for many institutions.

Consequently, TBMA-related instructional programs must be carefully designed to address variables beyond the control of participants, such as time constraints, work typologies, and degrees of abstraction. These considerations are essential to maintaining pedagogical integrity while enhancing instructional effectiveness. Addressing these challenges is critical to the continued integration and development of TBMA within contemporary art education.

1.1.2 Emotion Recognition Technologies in Contemporary Education

Advancements in machine learning have introduced emotion recognition technologies into educational contexts. These machine learning–based systems offer educators timely and sufficiently accurate feedback, enabling the implementation of adaptive instructional strategies that respond to students' emotional states. Such strategies have been shown to enhance engagement and learning outcomes (Abdullah and Alkan, 2022; Kanjo et al., 2015; Li et al., 2023; Palm et al., 2024).

Emotion recognition systems provide valuable insights into learners' emotional and cognitive states, thereby facilitating timely interventions that can significantly improve educational performance. Vistorte et al. conducted a systematic review of emotion recognition in education and emphasized the role of emotional feedback in enhancing student engagement (Vistorte et al., 2024). Their findings indicated that the integration of artificial intelligence–driven emotion recognition systems can significantly increase student motivation and academic performance, particularly in subjects that demand creativity and active participation. The incorporation of emotion recognition technologies in education has also been linked to the development of emotional intelligence. By providing feedback on students' emotional states, these systems promote self-awareness and empathy, traits that are essential for both artistic expression and collaboration in STEM disciplines (Barrett et al., 2019; Nandwani and Verma, 2021). Concurrently, the accessibility of smart mobile devices has contributed to a more flexible and self-paced learning environment (Lai and Zheng, 2018). The rise of digital technologies and the proliferation of smart mobile devices have enabled students to access a broad range of digital resources, collaborate in both physical and virtual spaces, and engage in interactive learning experiences (Crompton and Burke, 2018; Sharples et al., 2010).According to research by Tai et al., students who utilized mobile applications in art education demonstrated higher levels of engagement and motivation compared to those using traditional instructional methods (Tai et al., 2019). The convergence of mobile learning and convolutional neural networks (CNN) has introduced a technologically oriented approach to contemporary education. For instance, students can collect and analyze images using mobile devices, apply CNN models for classification or interpretation, and thereby gain deeper understanding of complex concepts (Wang et al., 2018). The application of CNN in education has been explored across several domains, including image recognition, handwriting analysis, and object detection, each of which contributes to the development of critical thinking and problem-solving skills (Chen et al., 2024). Moreover, the combination of CNN with mobile learning platforms has enabled personalized learning pathways tailored to students' individual performance and needs, thus supporting differentiated instruction (Ota et al., 2017).

Although the potential advantages of using smart mobile devices to support emotion recognition are evident, several challenges remain. These include accurately detecting facial expressions, ensuring secure data storage, addressing ethical concerns, and determining device appropriateness across educational settings. Furthermore, existing research has noted that the use of smart mobile devices may negatively affect students' attention spans and could lead to addictive behaviors (Sophonhiranrak, 2021). Educators may also require targeted training to effectively incorporate these technologies into their pedagogical practices. Addressing these challenges necessitates collaborative efforts among educators, technologists, and policymakers to design curriculum structures that are pedagogically sound and technologically appropriate.

1.2 Research Scope

This dissertation spans four interrelated domains: Time Based Media Art, medium theory in art education, emotion recognition and instructional design grounded in educational psychology. All domains converge on the central objective of designing and evaluating a curriculum that supports the structured appreciation of TBMA in higher education, with particular attention to how temporally organized viewing and guided reflection can facilitate emotional reappraisal and deeper meaning making. Each domain plays a distinct yet interdependent role in addressing the research objectives and constructing the conceptual framework of the study. Rather than operating as isolated topics, these domains are interconnected to support a psychologically informed learning experience that aligns with contemporary expectations for media arts education, where appreciation is understood as a developmental process involving perception, affect, interpretation, and value formation. Accordingly, this dissertation engages with a hybrid research field that integrates aesthetic experience theories with educational psychology and curriculum research, and proposes an instructional framework suited to the temporal and interpretive demands of TBMA.

The first core domain explored is TBMA, a contemporary art category characterized by temporal unfolding, sensory immersion, and performativity (Grau, 2010). TBMA has gained increasing prominence in both artistic practice and arts education, particularly in light of advancements in digital media and interdisciplinary approaches (Sun et al., 2023). From a curricular perspective, TBMA presents unique challenges due to its temporality, non linear narrative structures, and the emotional fluctuations it elicits from audiences. These characteristics demand instructional strategies that move beyond static observation or textual interpretation, requiring learners to engage emotionally and cognitively with artworks that unfold over time. At the same time, TBMA offers a distinctive opportunity for art appreciation education because learners' understanding, interest, and liking often develop through repeated encounters, perspective shifting, and reflective articulation, rather than through immediate recognition of subject matter or technique. Drawing on recent scholarship in curriculum time studies (Leek et al., 2024; Scheerens et al., 2013), this research conceptualizes instructional time not merely as a structural constraint but as an active pedagogical agent that influences the rhythm, sequence, and depth of TBMA engagement. Against this backdrop, the study investigates how structured time allocation and reflection cycles can support sustained attention, facilitate reappraisal, and promote interpretive growth in TBMA appreciation.

The second domain concerns the reconceptualization of the medium within TBMA. This part of the study draws from theoretical discourses in medium studies, art history, and aesthetics to examine how TBMA challenges traditional definitions of artistic medium (Krauss, 1999; Manovich, 2001), particularly those grounded in materiality, temporality, interactivity, and technological mediation. In TBMA, the medium is no longer a passive material conduit but a dynamic interface influenced by audience interaction, sensor input, and algorithmic transformation. This study extends the notion of medium by positioning emotion not only as an evaluative tool but also as a pedagogical medium that influences students' experience, interpretation, and reflection of digital artworks. From an art appreciation perspective, this reconceptualization clarifies why meaning making in TBMA is often inseparable from the conditions of encounter, including the pacing of attention, the sequencing of cues, and the learner's reflective distance from the experience. It also supports a curriculum logic in which medium awareness becomes part of appreciation, enabling students to articulate how temporal structure and technical mediation influence what the work communicates and how it is valued.

The third domain of this research focuses on facial emotion recognition (FER), particularly through the use of convolutional neural networks (CNN). FER has emerged as one of the most widely adopted modalities in affective computing due to its balance of accuracy, efficiency, and unobtrusiveness (Ko, 2018; Whitehill et al., 2014). Unlike voice or physiology based systems, FER can be implemented with standard cameras, making it especially suitable for classroom environments. CNN models significantly improve the precision of FER through automated feature extraction and real time classification of subtle facial expressions. Their modular and scalable design supports implementation on mobile devices such as tablets and smartphones, aligning with this study's emphasis on mobile supported learning environments (Alzubaidi et al., 2021; Canal et al., 2022). In this research, CNN based FER is used to capture spontaneous affective responses during TBMA exposure, providing time sensitive indicators that complement students' reflective accounts. However, the analytic function of these indicators is positioned in service of instructional design, supporting the examination of how appreciation unfolds across initial encounter and delayed reflection, rather than treating automated recognition as the focus of the inquiry. The study also addresses key concerns related to privacy (Tanwar et al., 2024), algorithmic bias, and educator trust, acknowledging the ethical implications of using artificial intelligence to interpret human emotion in educational contexts.

The research also incorporates text based emotion recognition through emotion dictionaries. While facial measures capture immediate and often involuntary expressions, textual analysis provides reflective and context sensitive insights into students' internal states. Drawing on research in natural language processing (Hutto and Gilbert, 2014; Mohammad, 2012), the study employs emotion dictionaries to classify students' written reflections based on emotional valence and category. This method is particularly well suited to educational settings, where textual reflection is a core mechanism of learning and formative assessment in art appreciation. In this research, textual self reports are interpreted as evidence of meaning making, including how students justify understanding, articulate interest, and explain liking or disliking in relation to the work's temporal structure and medium conditions. Textual accounts are triangulated with facial indicators to construct a more comprehensive picture of how affective experience is transformed into reflective judgment during TBMA learning.

By integrating theoretical and practical contributions across these four domains, this research addresses the practical challenge of designing instruction for time dynamic media art, proposes a curriculum structure that supports repeated encounter and guided reflection, and strengthens art appreciation research by modeling appreciation as a temporally extended process that integrates affect, interpretation, and evaluative judgment. It also contributes theoretical insights to medium theory by clarifying how temporality and technical mediation influence the conditions under which appreciation develops, thereby offering an integrated framework for curriculum design and evaluation in higher education TBMA contexts.

1.3 Research Objectives

The Main Research Objective (MRO) of this dissertation is to investigate whether and how emotional reappraisal in TBMA education generates effects on students' appreciation of TBMA works in Chinese higher art education. The dissertation focuses on both the feasibility of employing mobile based emotion recognition in higher education and the pedagogical value of structuring repeated viewing and reflection processes to support learning related outcomes.

Accordingly, this dissertation explores two sub objectives. The first sub objective (SRO1) is to examine the feasibility of employing emotion recognition in the art classroom context of Chinese universities, with particular attention to ethical concerns, device ownership, willingness to participate, and the perceived accuracy of assessment. At the same time, SRO1 explores how repeated viewings of TBMA works influence students' emotional reappraisal and whether such processes show individual level variations or converge toward collective trends. The second sub objective (SRO2) is to explore the effects of emotional reappraisal on students' appreciation of TBMA, with appreciation defined to include understanding, interest, and liking. This involves testing whether structured reappraisal contributes to deeper and more reflective engagement with TBMA artworks.

Furthermore, Figure 1.1 illustrates the relationship between the research objectives and the three studies conducted.

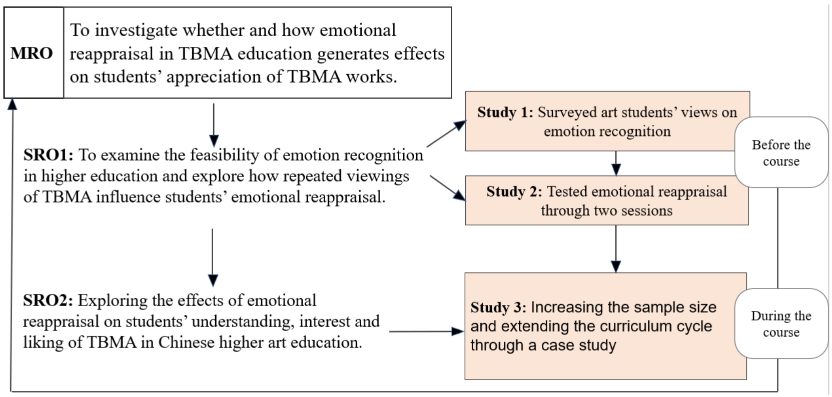

Figure 1.1: The relationship between research objectives and studies

As shown, to realize these objectives, the research is organized into three studies. Study 1 surveyed art students' views on emotion recognition and clarified the ethical and practical conditions necessary for implementation before the course. Study 2 tested emotional reappraisal through two sessions under these conditions, examining whether students' emotional change patterns were dispersed individually or aligned with collective trends before the course. Study 3 extended the investigation by increasing the sample size and lengthening the curriculum cycle through a case study during the course, which enabled the analysis of the effects of emotional reappraisal on understanding, interest, and liking. Together, these studies provide a sequential path from feasibility to empirical testing and then to curriculum level evaluation, thereby systematically addressing both SRO1 and SRO2 and contributing to the realization of the MRO.

1.4 Significance of this Research

This dissertation has conceptual, methodological, technological, and pedagogical significance by advancing an instructional design for TBMA appreciation that treats emotion as a learnable and revisable component of reflective meaning making in higher education. Rather than approaching emotion as an incidental reaction to media art, the study formalizes how temporally organized viewing and guided reflection can structure students' engagement with complex, nonlinear artworks and support deeper interpretive learning. Within technology enhanced classrooms where mobile devices are already normalized as learning tools, this dissertation proposes and empirically substantiates a multimodal, time structured framework for tracing students' developing affective responses as part of the learning process, with emotion related data serving as feedback for teaching decisions rather than as the endpoint of the research.

Methodologically and in terms of educational psychology, the dissertation contributes a sequential and evidence based pathway from feasibility to classroom validation and longitudinal evaluation. Study 1 clarifies the design constraints under which emotion related data collection can be instructionally acceptable, showing conditional acceptance influenced by perceived intrusiveness, privacy risk, and fairness concerns, and translating these concerns into privacy by design requirements such as transparent consent, meaningful opt out options, and learner agency through student owned devices. Building on these constraints, Study 2 shows that a structured interval design, supported by guided re viewing and reflection, can elicit emotional reappraisal by increasing the richness of non dominant emotional signals while keeping dominant emotional categories comparatively stable. This establishes a replicable protocol for capturing change processes that are central to reflective learning rather than relying on single shot impressions. Extending beyond short term effects, Study 3 embeds the same two time points, Initial Impression (II) and Delayed Impression (DI), within a four week curriculum cycle using diversified TBMA stimuli. The results show that delayed reflection consistently supports higher understanding than initial viewing, while also demonstrating that emotional change relates differently to understanding than it does to interest and liking. Across weeks, non dominant emotional indicators are the most stable predictors, with affective diversification and articulation supporting gains in understanding more consistently than gains in interest and liking, which appear to depend more heavily on reflective textual meaning making than on facial impressions. Collectively, these results yield actionable instructional design implications, including when to schedule reflection prompts, how to pace repeated encounters to support elaboration without fatigue, and how to interpret divergence between outward expression and self reported meaning in light of emotion regulation processes such as suppression and reappraisal.

The dissertation also strengthens its significance for Knowledge Science by showing that emotion plays an active role in art learning and continues to influence how understanding develops, and by describing emotional knowledge as changes that develop over time and can be observed through different forms of expression, rather than as fixed emotional labels. Across the three studies, the work connects affective dynamics to learning relevant outcomes such as understanding, interest, and liking, enabling a portable account of how temporal organization, guided attention, and reflective articulation co produce meaning in TBMA. This trajectory based perspective differentiates pathways that support explanation and interpretation from pathways that support preference formation, and it provides a Knowledge Science informed basis for designing instructional feedback loops in which educators can adjust sequencing, scaffolding, and reflection in response to students' developing meaning making.

1.5 Structure of the Dissertation

This dissertation is comprised of the following seven chapters:

Chapter 1 (Introduction) provides an overview of the research, which includes the background, research objectives, significance, and structure of the study.

Chapter 2 (Literature Review) describes the contextual framework of the study and critically reviews the literature related to TBMA, medium and emotion recognition in higher education. This chapter also provides the theoretical foundation for designing a TBMA course setting that incorporates emotion recognition technologies in the context of higher education.

Chapter 3 (Methodology) describes the comprehensive methodological framework employed across the three interconnected studies that structure this dissertation. Anchored in the overarching goal of exploring how emotion recognition technologies can enhance the appreciation of TBMA in higher education settings, each study is methodologically distinct yet conceptually interrelated. The structure of this chapter reflects the sequential and accumulative nature of the research design, beginning with a cross-sectional survey (Study 1), progressing to a controlled experimental design (Study 2), and culminating in a curriculum-integrated case study (Study 3).

Chapter 4 (Results of Study 1 and Study 2) reports the empirical results from the dissertation's first two phases. It first presents Study 1 survey findings, including instrument reliability and validity, students' device ownership and baseline attitudes toward educational use, and predictors and demographic differences in willingness to adopt emotion recognition in classroom contexts. It then presents Study 2 results from the mixed-methods experiment, focusing on consistency of dominant emotions between II and DI, and comparative analyses of II versus DI, including both between-group comparisons and changes in the diversity and intensity of non-dominant emotions.

Chapter 5 (Results of Study 3) extends the earlier feasibility and validation phases by reporting the empirical findings of Study 3, which investigates a curriculum-integrated, four-week TBMA experience supported by emotion recognition. It presents dominant-emotion consistency between II and DI, systematic differences in non-dominant emotions across the two time points, differential regression results linking emotional change to changes in art understanding, interest, and liking, and linear mixed model analyses that estimate time and emotion-related predictors of these outcomes across repeated weekly sessions.

Chapter 6 (Discussion and Conclusion) synthesizes findings across all three studies and situates them within TBMA education, educational technology, and Knowledge Science. It discusses Study 1 in terms of conditional acceptance, perceived intrusiveness, and ethical concerns, translating these into design and implementation implications for classroom emotion recognition. It interprets Study 2 as evidence for re-viewing and emotional reappraisal processes, distinguishing roles of negative versus positive appraisals and clarifying how mere exposure and conscious reappraisal may operate in TBMA contexts. It then integrates Study 3's longitudinal evidence, emphasizing how non-dominant emotion indicators relate to changes in understanding, interest, and liking, and uses the combined results to articulate curriculum-relevant design principles, limitations and future research directions, and the dissertation's contributions to Knowledge Science.

Chapter 2

Literature Review

2.1 Theoretical Foundation from Artistic Experience

Traditionally, aesthetic experience has been conceptualized as a moment of disinterested pleasure, characterized by cognitive harmony and detached contemplation (Osborne, 1979; Walton, 1993). Such a perspective reduces aesthetic experience to a moment of insight, overlooking the affective and cognitive transitions that often occur before, during, or after aesthetic perception. This linear model, which assumes a defined endpoint, fails to account for the emotional and cognitive shifts that precede insight or self-transformation. As aesthetic models continue to develop, aesthetic experience is increasingly understood not as a passive act of observation, but as an interaction involving perception, cognition, and affective responses that unfold over time.

Recent theoretical frameworks, particularly the model proposed by Leder and Nadal (2004; updated in 2014), offer a structured approach to understanding how individuals engage with artworks(as shown in Figure 2.1). Their model presents aesthetic experience as a continuous, multi-stage process that begins with perceptual analysis and culminates in aesthetic judgment and emotional response. The model incorporates both automatic processes (perceptual fluency, affective states) and deliberate processes (interpretation, evaluation). Emotional-cognitive development is described along this continuum, from primitive affective reactions to reflective appraisals influenced by context, expertise, and social discourse (Leder and Nadal, 2014).

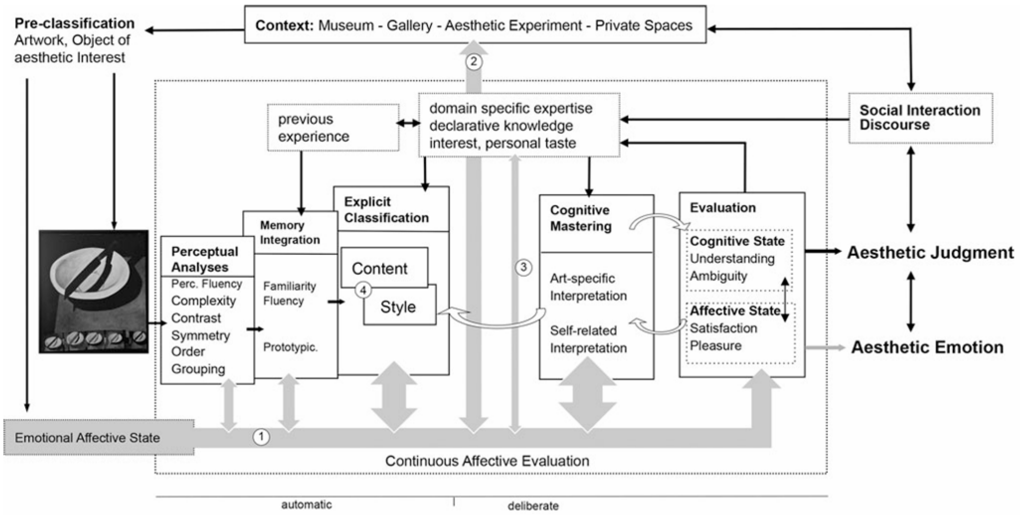

Figure 2.1: Model of aesthetic experiences (adapted from Leder, Belke, Oeberst and Augustin, 2004). Painting by Helmut Leder.

Leder et al.'s model explicitly distinguishes between perceptual-level affective states and aesthetic emotions arising from evaluative processes, such as awe or pleasure. The transition from basic to nuanced emotions reflects broader trajectories in cognitive and emotional regulation. Affective cognition develops from early childhood recognition of valence-based cues to the complex aesthetic emotions experienced in adulthood. According to Saarni's research (Saarni, 1999), emotional competence encompasses not only emotion recognition but also contextual understanding and expressive regulation, capacities that improve with age. While children may respond to symmetry or color due to perceptual fluency, adults are more likely to integrate prototypicality, style, and self-referential meanings, as reflected in the memory integration and cognitive mastering stages of the model. Keltner and Haidt define aesthetic emotions as those that connect individuals to broader meanings and transcendence (Keltner and Haidt, 2003), such as awe and wonder. These emotions become more intelligible through domain-specific expertise and social discourse, both of which are emphasized components in Leder et al.'s model. Supporting this view, studies by Silvia and Menninghaus et al. argue that aesthetic emotions are not instantaneous (Menninghaus et al., 2020; Silvia, 2005), but rather develop through repeated exposure, interpretation, and personal relevance. This aligns with the model's notion of continuous emotional evaluation, an developing process influenced by individual perspective and lifelong experience. Over time, exposure to diverse artworks fosters the development of declarative knowledge,

Table 2.1: The transformative model of Pelowski and Akiba.

Stage	Description
Stage 1: **Pre-encounter**	Involves cognitive framing and schema activation, but no artwork engagement.
Stage 2: **Cognitive Mastery**	May lead to initial success (pleasure) or failure (anxiety, threat).
Stage 3: **Secondary Control**	Triggered by unresolved conflict, producing tension, sadness, or avoidance.
Stage 4: **Meta-cognitive Assessment**	Enables reflection, schema-change, or transformation through emotional insight.
Stage 5: **Aesthetic Outcome**	Manifests in catharsis, relief, or epiphany.

enhances cognitive fluency, and refines emotional granularity (Barrett et al., 2001). These developments promote richer engagement with ambiguity and domain-specific interpretation, reinforcing the understanding that aesthetic appreciation is a continually developing achievement. Moreover, Pelowski and Akiba challenge the traditional notion of aesthetic experience as a linear endpoint (Pelowski and Akiba, 2011), proposing instead a five-stage model that emphasizes disruptions and reevaluation within aesthetic engagement(as shown in Table 2.1). Within this framework, affective cognition is understood not as a static appraisal but as a dynamic developmental process. These transformations correspond with increased emotional differentiation and regulatory capacity (Barrett et al., 2001; Saarni, 1999), reflecting broader psychological development.

Crucially, the advancement of aesthetic models extends beyond the understanding of artworks to include the reconstruction of self-image, the redefinition of values, and the attainment of deeper insight. Thus, affective cognition in aesthetic experience is not fixed but develops with cognitive maturity, experiential depth, and sociocultural context. Understanding aesthetic experience as a developmental affective journey, rather than a cognitive endpoint, offers a transformative perspective for art education. Students, especially within the time constraints of classroom environments, frequently encounter the kinds of disruption and reevaluation described by Pelowski. Facilitating this developmental journey through guided interpretation, reflective engagement, or structured curricula may promote deeper involvement with art. Accordingly, empirical research in this area calls for a longitudinal and process-oriented approach. Aesthetic emotions should not be tracked as isolated responses but rather as developing states that unfold across time and context.

2.2 Time-Based Media Art in Course Setting

2.2.1 From Media Art to Time-Based Media Art

Media art, also known as new media art, represents an intersection of art, technology, and communication. It encompasses artworks that are created, modified, or transmitted through digital technologies, with an emphasis on interactivity, processuality, and real-time participation (Grau, 2010). Unlike traditional art forms, media art is characterized by its reliance on technological media, fostering new modes of artistic expression and audience engagement.

The origins of media art can be traced back to the avant-garde movements of the early twentieth century, which experimented with the fusion of technology and art. The 1960s and 1970s witnessed the emergence of dynamic and interactive forms of art, laying the groundwork for contemporary media art (Carroll, 1985; Grau, 2010). Its development has been marked by rapid technological advancement and interdisciplinary collaboration. The advent of the internet and digital technologies in the late twentieth century significantly expanded the possibilities for artistic creation and dissemination. Artists began to explore virtual reality, interactive installations, and net art, thereby challenging conventional notions of authorship and audience participation (Hölling, 2017; Quaranta, 2013).

With the continued expansion of media practices, time-based media art (TBMA) has increasingly gained attention. TBMA refers to artworks that unfold over time, integrating elements such as video, film, audio, performance, and digital technologies. Unlike static art forms, TBMA engages audiences through temporal experience, thereby challenging conventional notions of conservation, documentation, and exhibition (Guggenheim Museum). Due to its dynamic nature, TBMA necessitates interdisciplinary approaches that bring together art history, media studies, conservation science, and digital technology (Ciobanu and Fernaeus, 2024; Peppler, 2010).

The emergence of TBMA in the 1960s and 1970s marked a significant shift in contemporary artistic practice. Artists began experimenting with video, performance, and digital media to create immersive experiences (Meigh-Andrews, 2013). The availability of affordable video equipment made media production more common and led to a surge in experimental artistic production. Over time, TBMA expanded to include interactive installations, virtual reality, and internet-based art, reflecting both technological advances and shifting audience expectations. Key characteristics of TBMA include its temporal dimension, interactivity, and reliance on technology (Laurenson, 2013). Such works often require specific playback devices and environmental conditions to function as intended. The obsolescence of technology and environments has rendered many works unexhibitable, posing significant challenges to their conservation (Phillips, 2012). Moreover, the variability in presentation, due to audience interaction and site-specific factors, further complicates documentation and preservation. As hardware and software continue to become obsolete, maintaining the functional integrity of these artworks grows increasingly difficult. Institutions are thus faced with critical decisions regarding whether to preserve original equipment, migrate content to newer formats, or emulate obsolete environments. Each approach has implications for the authenticity and integrity of the artwork (Vassos et al., 2016). Researchers involved in TBMA conservation must grapple with ethical dilemmas, as alterations or replacements of components may affect the authenticity of the work. Where possible, understanding the artist's intent and engaging directly with the artist allow for informed modifications within acceptable bounds, thereby ensuring that future presentations remain as faithful to the original as possible (Rinehart and Ippolito, 2022).

Currently, two mainstream strategies are employed in the preservation and re-presentation of TBMA: emulation and migration. Emulation involves recreating the original technological environment using modern tools, while migration refers to updating content to run on contemporary platforms. Both strategies aim to balance preservation and authenticity by offering interpretative frameworks for how TBMA may be exhibited when its original environment is no longer accessible or functional. Although controversial, especially when reinterpretation involves creative reconfiguration based on perceived artistic intent, such practices may extend the lifespan and relevance of time-based media art in the face of inevitable technological change.

Recognizing the potential of TBMA within educational contexts, numerous academic institutions have incorporated specialized programs aimed at training artists, conservators, and researchers in this developing field (see Table 2.2).

Table 2.2: Educational programs of TBMA in the United States, Japan and China.

University and Institution	Program Overview
New York University	Program: Conservation of Time-Based Media Art Survey of time-based media technologies and care; advanced training Interdisciplinary: CS, engineering, and film and video preservation Method: Collaboration with specialists and professional networks
University of Pennsylvania	Program: Time-Based and Interactive Media Focus: Moving image, digital technologies, and interactivity Outcome: Engage with emerging tools/methods and integrate into art/design practice
Sam Fox School of Design and Visual Arts	Program: Time-Based Media Art Focus: Time as the primary expressive medium Courses: Expanded cinema, sound environments, and media performance Emphasis: digital, electronic, acoustic, cinematic, and performative approaches
Maricopa Community Colleges	Program: Time-Based Media Focus: Production techniques, compositing, motion graphics, and interactive design Competencies: Media and photographic technologies, production equipment, and industry practices
University of Tennessee, Knoxville	Program: Time-Based art Focus: Animation, video, film, performance, installation, sound art, and interactive time-based media

Table 2.2 (continued)

Kyoto Seika University	Programs: Video and Media Art Focus: Artistic animation, short films, media arts, music videos, commercials, video installations, and projection mapping Emphasis: Creative expression and communication
Tama Art University	Programs: Art and Media Focus: Innovative expression in digital art and creative practice, video and animation production, programming, and 3D modeling Emphasis: Digital fabrication
Communication University of China	Program: Digital Media Art Focus: Ontological theory and industrial applications, intelligent media, interactive art, motion graphics, virtual societies, and digital entertainment Emphasis: A global perspective and a deep grounding in Chinese cultural heritage
Beijing Institute of Graphic Communication	Program: Digital Media Art Focus: Appreciation of digital media art, appreciation of video and animation art, and introduction to film aesthetics

While the term time-based media has been widely adopted in the United States, it is more commonly subsumed under broader categories such as media art, new media art, or digital media art in China and Japan.Institutions such as New York University and the University of Pennsylvania have established dedicated departments and certificate programs that explicitly label their curricula as time-based media, reflecting an epistemological and practical effort to articulate temporality as a central dimension of contemporary artistic production. Furthermore, American museums such as the Museum of Modern Art and the Guggenheim Museum have reinforced the academic relevance of the term through targeted initiatives in acquisition, conservation, and exhibition.

In contrast, the absence of the term time-based in Chinese and Japanese contexts does not necessarily indicate a conceptual oversight. Rather, it reflects different cultural traditions and linguistic frameworks. Time-based practices have been integrated into educational and artistic structures through alternative terminologies. For example, Kyoto Seika University and Tama Art University in Japan offer programs such as video and media art or art and media, which emphasize digital production, animation, and interactivity without using time-based media as a formal academic label. Similarly, institutions in China, such as the Communication University of China and the Beijing Institute of Graphic Communication, provide programs under the title of Digital Media Art, with a strong emphasis on technical training and industry application. Differences in museological and conservation practices have also contributed to uneven institutional support. While Western institutions have developed detailed preservation protocols for time-based media (Sherring, 2020), including strategies such as migration, emulation, and reinterpretation, museums in East Asia have tended to focus more on the exhibition of contemporary works rather than on documenting their technological dependencies.

Future international collaboration and curriculum development may benefit from recognizing these differences while seeking to integrate the conceptual rigor of time-based frameworks with region-specific pedagogical traditions.

2.2.2 Rethinking the Medium in TBMA

The concept of medium, once confined to the physical materials of art, has expanded into a complex, interrelated system of modalities encompassing temporal, experiential, and technological dimensions, functioning not merely as a channel for transmission but as an active agent in meaning-making. This expanded conception of the medium plays a constitutive role in how communication and meaning are formed in both everyday life and artistic expression. A medium constitutes the overall structure through which content is mediated, realized through material and perceptual means (Elleström, 2010). It consists of four essential modalities: material (physical matter), sensorial (sense perception), spatiotemporal (organization in space and time), and semiotic (systems of signs and meaning). Each modality contains specific forms, known as modes,for example, the visual mode within the sensorial modality. A comprehensive understanding of any medium requires recognition of all four modalities. In this framework, technical medium are viewed as the material realization of the latent properties inherent in a medium.

The concept of medium underwent a significant shift in the 1960s. Art critic Clement Greenberg advocated for the principle of medium specificity, asserting that each art form should exploit the unique qualities inherent to its medium (Greenberg, 1966). His modernist perspective emphasized the purity of artistic practice, for instance, painting was to focus on flatness and color rather than on illusory depth. This approach sought to preserve the integrity of individual medium by resisting interdisciplinary influence. However, the emergence of TBMA, including video, film, and digital installations, challenged the boundaries of medium specificity. Artists began integrating multiple medium, temporal elements, and interactive components, thereby complicating the notion of a singular, pure medium. This transition required a re-evaluation of the medium concept to address the complexity of contemporary artistic practices (Potts, 2004). Marshall McLuhan's assertion that "the medium is the message" highlighted the idea that the medium itself influences the content and perception of art (McLuhan, 2017). This concept resonated strongly with TBMA, where temporal and interactive features are integral to the meaning of the work.

Niklas Luhmann's system theory conceptualized medium as highly dissolvable structures: unordered aggregates that gain meaning only through forms imposed upon them. In art, the medium functions not merely as material but as a condition for the emergence of form (Luhmann and Roberts, 1987). Luhmann framed the medium and form in a dialectical relationship through which social meaning and communication are organized. Artistic works do not simply use medium; rather, they create medium by selecting, structuring, and formalizing elements from a field of conditions. In this sense, painting, music, or literature are not preexisting containers but dynamic systems organized through aesthetic operations in response to cultural expectations.

Rosalind Krauss introduced the concept of the post-medium condition to describe the contemporary artistic landscape that transcends traditional medium boundaries (Krauss, 1999). Krauss argued that medium is no longer defined by material substrate but by a set of conventions and practices. This view recognizes the fluidity of medium in contemporary art, where artists employ diverse technologies and processes to convey meaning. The post-medium condition is particularly relevant to TBMA, where the integration of time, motion, and interactivity disrupts conventional categorizations. Artists such as Nam June Paik and Bill Viola exemplify this approach, creating immersive experiences by fusing video, performance, and installation, thereby challenging traditional definitions of medium. The proliferation of digital technologies has further transformed the concept of medium in art. Manovich identified features such as variability, modularity, and automation as defining characteristics of media art, redefining medium as a dynamic system rather than a static material (Manovich, 2001). This evolution reflects the growing influence of software and algorithms in artistic production, positioning the medium as a conduit for interaction and real-time processing. Hansen expanded this perspective by emphasizing the experiential dimension of digital media. He posited that the medium is constituted through the specific interaction between viewer and artwork, underscoring the importance of perception and participation in defining medium (Hansen and Hansen, 2004). Bruno Latour offered an alternative framework, conceptualizing medium as networks of relations among human and non-human actors, including technologies, institutions, and audiences (Latour, 2005). This perspective highlights the interconnectedness of various elements in defining a medium, aligning closely with the multifaceted nature of TBMA.

Rancière critically reframed the notion of medium by introducing the concept of neutrality (Rancière, 2011). Rather than viewing the medium as a passive conduit or mere material, Rancière emphasized the tension between medium as a neutral tool and medium as a sensory condition. Thus, neutrality binds the technological apparatus, artistic vision, and sensory engagement within a unified experiential field. Alessio Chierico revisited the concept of medium specificity in the post-media era, noting how Greenberg's formalism and Krauss's critique continue to influence the theoretical lineage of new media art (Chierico, 2016). The perspectives of artists and scholars contribute to a nuanced understanding of medium that transcends material determinism. The concept of medium has shifted from formalist constraints to a post-media condition in which artists and researchers from other fields co-invent, remix, and critique the medium itself.

Through the above conceptual review and historical development of TBMA and the notion of medium, several critical gaps are identified, particularly in the context of TBMA and transformative aesthetic experiences, alongside rich opportunities for theoretical advancement and practical application. While traditional medium theories have emphasized structural or material foundations, they often overlook the temporal and processual dimensions that are central to TBMA and to models of affective-cognitive transformation in artistic experience. The failure to address how the medium as a temporal and experiential structure unfolds over time and interacts with audiences' developing emotional and cognitive processes, from confusion to reflection and insight, has limited the explanatory power of these frameworks. Moreover, although theorists such as Manovich, Hansen, and Latour have made significant contributions, the role of technological mediation remains under-theorized, particularly regarding how digital infrastructures influence aesthetic experiences from creation to preservation. At the same time, these gaps present valuable opportunities: reconceptualizing the medium as a dynamic interface between technology, artist, and audience; integrating models of aesthetic experience into the preservation and interpretation of TBMA; and designing art curricula that connect theory, practice, and audience engagement.

2.3 Recognition of Facial Emotions and Textual Emotions

2.3.1 Integration of Facial and Textual Emotion Recognition Technologies in Education

Emotion recognition represented a multifaceted domain within affective computing, encompassing various modalities such as text analysis, facial expressions, vocal intonation, and physiological signals. During the past two decades, emotion recognition technology has progressed from a conceptual framework to a practical tool within educational settings. Initially rooted in the field of affective computing, the idea of enabling computers to detect and respond to human emotions was introduced by Picard (Picard, 2000). Picard's foundational work proposed that when emotional intelligence is embedded into computational systems, it can enhance human-computer interactions in various domains, including education.

Early applications primarily focused on unimodal approaches, such as facial expression analysis and vocal emotion detection (Cowie and Cornelius, 2003; Pantic and Rothkrantz, 2000). These studies demonstrated that emotional cues could be extracted from observable behaviors, such as facial muscle movements, vocal prosody, and posture, in order to infer learners' emotional states. At this stage, applications remained largely experimental and were restricted to controlled environments.As computational power and sensor technologies advanced, research shifted toward multimodal systems. For example, emotion recognition systems began integrating facial, vocal, and physiological data to identify emotional responses to video stimuli (Tse et al., 2002). This multimodal approach enabled more accurate and reliable emotion classification, thereby allowing real-time adaptation of educational content. Similarly, D'Mello and Graesser introduced a model capable of detecting both cognitive and affective states and of providing personalized feedback based on students' levels of engagement and confusion (D'Mello and Graesser, 2012).The integration of text-based recognition with facial expression recognition was shown to produce more robust and accurate emotion recognition systems. By combining linguistic and visual data, these multimodal systems addressed the inherent limitations of each individual modality. For instance, ambiguous textual expressions could be clarified through corresponding facial cues, while subtle facial expressions could be better interpreted when contextualized with accompanying text (Poria et al., 2017).

In recent years, particularly since 2020, there has been a notable increase in the development and implementation of emotion recognition systems in both physical and virtual classrooms. Llumba et al. conducted a scoping review of real-time emotion detection technologies, highlighting their potential to optimize classroom dynamics and improve instructional effectiveness (Gupta et al., 2023). These technologies have been increasingly adopted in various educational contexts, including intelligent tutoring systems, e-learning dashboards, classroom engagement analytics, and social-emotional learning programs. Notably, Katirai explored the intersection of emotion recognition and educational sustainability, underscoring its potential to identify and support vulnerable learners, thus promoting equity and fairness in education (Katirai, 2024). The application of emotion recognition technology has extended beyond traditional classrooms to include online learning platforms. Emotion recognition systems were integrated into these platforms to assess student engagement, detect confusion or frustration, and provide timely interventions, thereby simulating the responsiveness of face-to-face interactions. The integration of facial expression recognition with textual emotion recognition enhanced the overall reliability of emotion detection and contributed to a more comprehensive understanding of human affect. Franěk et al. examined emotional responses to natural scenes by analyzing participants' facial expressions using automated recognition software alongside their self-reported emotional states. Correlations were observed between objective facial data and subjective reports, indicating that combining these approaches could lead to a more accurate representation of emotional responses (Franěk et al., 2022). This integrative method proved particularly valuable in domains such as education, psychology, and human and computer interaction, where understanding the multifaceted nature of emotion was considered essential.

Despite these advancements, the widespread adoption of emotion recognition technologies in art education continues to face significant challenges. Concerns have been raised regarding privacy, data security, and the ethical implications of monitoring students' emotional states. Moreover, the accuracy of emotion detection algorithms and their generalizability across diverse populations remain areas requiring further investigation. The lack of appropriate equipment and technical expertise within many arts education institutions further complicates implementation. Addressing these issues is critical to realizing the full potential of emotion recognition technologies in fostering inclusive and effective arts education experiences. In response to these challenges, the present research further investigates recognition technologies based on facial expressions and self-reported text, aiming to identify a hybrid approach that is easy to implement, precise, rapid, and efficient within the context of structured TBMA curricula.

2.3.2 Facial Emotion Recognition Based on CNN

Facial expression recognition (FER) emerged as a dominant modality in emotion recognition systems due to its balance between accuracy, efficiency, and noninvasiveness. Unlike voice based or physiology based approaches, FER relied on visual cues derived from facial movements, which were generally understood across diverse cultural contexts (Ekman and Friesen, 1971; Ekman et al., 1987). This universality enhanced the generalizability of FER systems, making them suitable for heterogeneous educational populations. In contrast, voice based systems were susceptible to linguistic variation and environmental noise, while physiological sensing typically required intrusive hardware (Picard and Healey, 1997; Schuller et al., 2011). FER could be implemented using standard cameras, enabling real time analysis of spontaneous expressions with minimal disruption to learners (Ko, 2018; Monkaresi et al., 2016).

In educational contexts, particularly within classroom settings, FER offered several pedagogical advantages. By interpreting learners' facial cues, FER systems enabled the monitoring of engagement, detection of confusion, and facilitation of adaptive instruction (Whitehill et al., 2014). These systems enhanced teachers' situational awareness, allowing timely interventions that improved learning outcomes and classroom dynamics. FER was also employed to personalize the learning experience by adjusting content delivery in response to students' emotional states and demonstrated effectiveness in remote learning by compensating for the lack of physical presence (Bosch et al., 2016). FER aligned more closely with the goals of affective computing, to develop scalable, non-intrusive, and context-sensitive emotion-aware technologies that respond to real-world educational dynamics (D'mello and Kory, 2015).

Among various emotion recognition techniques, FER based on convolutional neural networks (CNN) has become one of the most effective and widely adopted approaches (Alzubaidi et al., 2021; Canal et al., 2022). CNN, as a class of deep learning algorithms, are particularly well-suited for image-based tasks, such as detecting subtle variations in facial markers that correspond to emotional states. Early FER systems required manual feature extraction, a process that heavily depended on domain-specific expertise and labor-intensive fine-tuning. In contrast, CNN offered automated feature learning capabilities, which significantly simplified system development. Following breakthroughs in image classification, the application of CNN to FER began to receive increasing attention after 2010. Unlike conventional models that relied on handcrafted features, CNN were capable of extracting features directly from raw pixel data, thereby substantially improving the classification accuracy of FER systems (Krizhevsky et al., 2012). CNN-based FER systems typically required minimal user calibration, which enhanced their usability across diverse learner populations (Sheng and Lau, 2024; Singh et al., 2023). This characteristic contributed to broader acceptance among educators and students by minimizing setup time and reducing cognitive load.

Moreover, the versatility of CNN enabled their integration into resource-constrained devices such as tablets and smartphones, which have been increasingly utilized in hybrid and remote learning environments. This implementation flexibility was considered crucial for scalable deployment across both traditional and digital classrooms. For instance, even relatively simple CNN architectures were capable of classifying facial expressions into basic emotional categories with minimal preprocessing, allowing their integration into lightweight educational applications (Hadiprakoso et al., 2020; Küçük et al., 2019).

CNN also demonstrated strong performance in detection accuracy. Their ability to learn hierarchical representations enabled the capture of complex spatial features and subtle facial muscle movements. CNN-based models outperformed traditional classifiers in recognizing facial expressions across diverse datasets, achieving high levels of precision and recall. When trained on large, annotated datasets such as FER2013 and CK+, these models typically achieved accuracy rates exceeding 90 percent (Alzubaidi et al., 2021; Ben Fredj et al., 2021). Such a high level of accuracy was regarded as critical for educational applications, as incorrect emotion detection could result in misinterpretation of learner engagement and hinder the effectiveness of instructional interventions. Therefore, the increasing reliability of CNN significantly enhanced the pedagogical value of FER in adaptive learning technologies.

These studies highlighted the superior performance of CNN over classical methods, particularly when trained on large and diverse datasets. From a technical standpoint, explainability, teacher trust, and model transparency have been recognized as critical factors for the acceptance of AI-driven systems in education.

2.3.3 Textual Emotion Recognition Based on Emotion Dictionary

In the field of natural language processing (NLP), sentiment recognition and emotion recognition in text were identified as two key yet distinct tasks. Sentiment recognition primarily focused on determining the overall polarity of a text, typically classifying it as positive, negative, or neutral. This approach offered a general overview of the text's tone, making it particularly useful in applications such as product reviews or public opinion analysis (Pang et al., 2008). In contrast, emotion recognition aimed at identifying specific emotional states expressed within the text, such as joy, anger, sadness, or fear. This level of granularity provided a more nuanced understanding of the underlying affective signals, which was critical in contexts such as mental health assessment or empathetic human–machine interaction (Mohammad, 2012).

The methodologies employed in these tasks further highlighted the distinctions between sentiment recognition and emotion recognition. Emotion analysis often utilized lexicon-based approaches or machine learning models trained on labeled datasets to classify texts according to their emotional polarity. For example, tools such as the Valence Aware Dictionary and Sentiment Reasoner were designed to analyze emotional content in social media texts, accounting for factors such as capitalization and punctuation (Hutto and Gilbert, 2014). Emotion recognition, on the other hand, typically required more sophisticated models capable of capturing subtle affective cues. Recent advancements incorporated deep learning architectures, such as bidirectional long-short-term memory networks and transformer-based models, to enhance the accuracy of emotion detection in textual data (Zhou et al., 2018).The applications of these analytical techniques were diverse and context-dependent. Emotion recognition proved particularly useful in more personalized domains, such as tailoring educational content based on students' emotional states or providing support in mental health interventions (Calvo and D'Mello, 2010). While sentiment recognition and emotion recognition were closely related, they served distinct purposes within NLP. Sentiment analysis broadly categorized texts based on their overall tone, whereas emotion recognition offered detailed insights into specific emotional expressions.

The emotion dictionary, also known as affective dictionaries, referred to curated lists of words associated with predefined emotional scores and played a crucial role in the automated evaluation of textual emotion. Within educational contexts, these lexicons were instrumental in analyzing emotional content in textual data such as student feedback, online forums, and course evaluations (Vinodhini and Chandrasekaran, 2012). One of the primary advantages of emotion dictionaries was their provision of a standardized and interpretable framework for emotion analysis, which allowed for consistent evaluation across diverse datasets. A notable strength of lexicon-based emotion analysis in education was its efficiency in processing large volumes of textual data. Educational institutions often collected extensive qualitative feedback from students, and manual analysis could be time-consuming. Lexicon-based methods enabled rapid processing of these data, offering timely insight into student perceptions and experiences. For instance, Misuraca et al. employed emotion analysis tools to evaluate student feedback, demonstrating the practical utility of lexicon-based approaches in educational settings (Misuraca et al., 2020). Moreover, emotion dictionaries could be customized for specific educational domains, thereby enhancing their contextual relevance and accuracy. General-purpose lexicons might fail to capture the nuanced meanings of terms commonly used in educational discourse, where certain words could carry context-specific connotations. By developing domain-specific dictionaries, researchers were able to improve the precision of emotion analysis in educational texts (Toba et al., 2024). The integration of emotion dictionaries into educational analytics also supported the identification of temporal trends and patterns in student feedback (Shaik et al., 2023). Through systematic analysis of emotions expressed in course evaluations, educators were better positioned to monitor shifts in student attitudes and to adjust their instructional strategies accordingly.

Based on a review of current developments in the field of emotion recognition, this research integrates facial emotion and text-based self-reporting into a structured, longitudinal curriculum for the appreciation of TBMA, providing a powerful and complementary approach to understanding students' emotional responses. Facial recognition captures real-time, involuntary reactions, such as surprise, confusion, or engagement, providing immediate physiological cues in response to dynamic media stimuli. In contrast, text-based self-reports offer reflective, nuanced, and culturally contextualized insights into students' internal emotional and cognitive states. The integration of these two modalities compensates for the limitations inherent in each. Facial data enhance the spontaneity and objectivity of subjective reports, while textual inputs clarify ambiguous expressions and reveal deeper layers of meaning. This multimodal approach increases both the accuracy and depth of affective analysis, thereby supporting adaptive instructional strategies and enabling longitudinal tracking of emotional development. Such integration is particularly valuable for capturing the developing affective dimensions of aesthetic experience in TBMA contexts.

Chapter 3

Methodology

3.1 Research Method of Study 1

3.1.1 An Overview of Study 1

Study 1 adopted a combination of qualitative and quantitative methodology, focusing specifically on undergraduate students' perceptions of utilizing smart mobile devices to capture facial expressions. A self-developed questionnaire was used to gather data from participants across five dimensions: basic demographic information, willingness to use smart mobile devices for learning purposes, willingness to use smartphones for capturing facial expressions, level of ethical concerns and participants' willingness after being presented with proposed solutions to these concerns. To fulfill the objectives of this research, two research questions were formulated:

RQ1: What is the prevalence of smart mobile device ownership among participants, and what are their general attitudes toward the use of such devices in educational settings?

RQ2: To what extent is willingness to use smart mobile devices to capture facial expressions explained by ethical concerns and by willingness to use mobile devices for learning purposes, and does this willingness differ by gender and academic year?

3.1.2 Participants

A valid sample comprising 199 undergraduate students from Dalian Polytechnic University and Shenyang Normal University participated in this study. Participants ranged in age from 19 to 24 years ($M_{\text{age}} = 20.9$; $SD_{\text{age}} = 1.17$), with females accounting for 71.4% (n = 142) of the sample. Regarding the year of study, sophomores constituted 35.7% (n = 71), juniors 46.2% (n = 92), and seniors 18.1% (n = 36). Freshmen were excluded from the study due to their involvement in foundational art courses. The confidentiality and anonymity of participants were maintained throughout the research process, and data were collected on-site from all participants. Prior to participation, informed consent forms were obtained, ensuring protection of participants' dignity and rights. All participants were fully informed about the objectives and procedures of the research, including details concerning data collection and usage. Additionally, all informed consent procedures were reviewed and approved by the deans of Dalian Polytechnic University and Shenyang Normal University, thus ensuring compliance with institutional and ethical guidelines.

3.1.3 Instrument Development

The capture of students' facial expressions has become increasingly prevalent in educational research as an indicator of engagement, comprehension, and emotional response (Birnhack and Perry-Hazan, 2020; Pabba and Kumar, 2022). Compared to traditional fixed equipment or desktop setups, smartphones offer several distinct advantages within classroom environments where flexibility, unobtrusiveness, and ethical considerations are critical. The questionnaire developed for this study was designed based on these investigated advantages, with the aim of assessing students' perceptions of smartphone-based facial expression recording (Al Fakih et al., 2020; Chen and Yan, 2016; Ekanayake and Wishart, 2014). The table (Table 3.1) summarizes the key benefits of using smartphones in this context, particularly in relation to educational data collection and emotion research.

Table 3.1: Advantages of using smartphones for capturing facial expressions in educational settings.

Aspect	Advantage
Unobtrusiveness	Smartphones are less intrusive and can be discreetly placed, reducing the observer effect and maintaining natural student behavior.
Contextual Integration	Smartphones could be embedded into classroom settings (used as timers, dictionaries, or classroom tools), making their presence less obvious and more accepted by students.
Flexibility of Angle and Proximity	Smartphones could be easily positioned at student desk level or near eye-line, providing better angles for facial recognition without bulky equipment.

Table 3.1 (continued)

Rapid Deployment	Unlike fixed recording systems, smartphones allow researchers or teachers to quickly set up and begin recording at any location within seconds.
Ethical Customization	Consent protocols can be integrated into app-based systems on smartphones, allowing for real-time, participant-controlled recording settings or opt-out functionality.
Localized Data Capture	Smartphones enable one-on-one or small-group recording without affecting the entire classroom, offering more precise targeting for micro-level interaction analysis.
Immediate Data Review	Educators or researchers can instantly review footage after class to reflect on student reactions, enabling quicker iteration in teaching or research design.

Based on the aforementioned investigation, to explore how students perceive and respond to such smartphone-based methods in actual classroom settings, a basic instrument was developed for this study (Table 3.2). The questionnaire is 5-point Likert scale and the scale ranged from 1 (strongly disagree) to 5 (strongly agree).

Table 3.2: Instrument for assessing student acceptance of facial expression recording using smart mobile devices.

Question No.	Item
Q1	What is your gender? a. Male b. Female
Q2	What is your age?
Q3	What year are you currently in at university? a. 1st year (Freshman) b. 2nd year (Sophomore) c. 3rd year (Junior) d. 4th year (Senior)
Q4	Do you own a smartphone or other digital mobile devices capable of recording photos or videos? a. Yes b. No

Table 3.2 (continued)

Q5	I regularly use smart mobile devices for academic purposes.
Q6	I believe that using smart mobile devices in educational projects is beneficial.
Q7	I support allowing the use of smart mobile devices to record photos or videos for educational projects.
Q8	I have positive expectations about using smart mobile devices to record my facial expressions for educational projects.
Q9	I believe that using smart mobile devices to record my facial expressions enhances my learning experience.
Q10	I believe that analyzing facial expressions can accurately reflect my emotional states.
Q11	I would notice a change in my concentration if I were aware that my facial expressions were being recorded during class.
Q12	I feel comfortable being recorded during class for the purpose of analyzing my facial expressions.
Q13	I believe that analyzing facial expressions could lead to bias or misinterpretation in educational assessments.
Q14	I am concerned about the ethical implications of analyzing facial expressions for educational purposes.
Q15	I am concerned that using smart mobile devices for facial expression recording in education may lead to fairness issues.

Table 3.2 (continued)

Q16	Which of the following do you think would improve your expectations and perceptions of your facial expressions being captured on a mobile device? 1. I prefer to use my own smartphone or personal device. 2. I prefer recordings to be limited to a short period of time rather than continuously during class. 3. I feel more comfortable if the device includes a beautification function. 4. I would like to receive feedback opportunities where I can express concerns or make suggestions. 5. I want a clear explanation of the purpose, benefits, and scope of data collection, along with informed consent. 6. I want the option to voluntarily opt out at any point during the project. 7. None of the above.
Q17	If the conditions I selected in Question 16 were fully considered, I would be willing to accept the use of smart mobile devices to record facial expressions in educational programs.

In addition, a multiple-choice question was also included. This questionnaire aims to gather initial feedback on students' attitudes, expectations, and concerns regarding the use of smart mobile devices for recording facial expressions during educational activities. As this was an foundational investigation, the instrument was intentionally designed to be simple and user-friendly, rather than relying on complex measurement models or psychological constructs.

3.1.4 Analysis Procedure

To address RO1, descriptive statistics, specifically frequency counts and percentages, were employed to quantify the proportion of participants reporting ownership of smart mobile devices, such as smartphones or tablets. The primary analytical goal was to establish a baseline criterion for inclusion in further analysis, based on the assumption that device ownership is a necessary prerequisite for meaningful participation in the study topic. Frequency analysis, a fundamental method for summarizing categorical data, was used here to determine whether the sample was suitable for exploring deeper attitudinal responses toward mobile-based educational interventions.

The study utilized a combination of descriptive statistics (means and standard deviations) and bar chart visualizations to summarize participants' responses across four attitudinal constructs: (1) general acceptance of smart mobile devices, (2) acceptance of facial expression recording, (3) ethical and fairness concerns, and (4) conditional acceptance based on ethical safeguards. These dimensions were grounded in prior literature and validated through reliability analysis and exploratory factor analysis. Descriptive statistics provided an empirical basis for identifying relative attitudinal patterns, such as whether general acceptance was higher or lower than conditional acceptance, and whether ethical concerns emerged as salient across the sample. These measures were essential for describing central tendencies and variances, offering insight into the distributional characteristics of each construct and informing interpretations in subsequent inferential analyses. Additionally, responses to the multiple-choice item (Q16), which captured participants' preferred ethical conditions for accepting facial expression recording, were visualized using bar charts. The objective was to identify which conditions, such as informed consent, device control, or beautification features, were most frequently endorsed, thereby offering a preliminary understanding of the ethical framework influencing students' conditional acceptance, as measured in Q17.

To address RO2, this study used a two-step inferential analysis strategy. First, Pearson product-moment correlation was used to assess the bivariate linear relationships among three continuous variables: general acceptance, acceptance of facial expression recording, and ethical concerns. Provided that the assumption of normality was reasonably met, this method was appropriate for examining the strength and direction of associations between scale-based variables derived from Likert items. The resulting correlation matrix informed the extent to which higher levels of general acceptance or ethical awareness were associated with higher or lower levels of acceptance of facial expression recording. Subsequently, to evaluate the predictive power of general acceptance and ethical concerns on the outcome variable, acceptance of facial expression recording, a multiple linear regression analysis was conducted. This analysis determined the unique and combined contributions of the independent variables in explaining the variance in the dependent variable. The regression coefficients, both unstandardized and standardized, indicated the magnitude and direction of the effects, while significance tests clarified whether the predictors were statistically meaningful. The R^2 value quantified the overall explanatory power of the model. This procedure represented a standard approach for testing hypotheses about causal relationships between continuous predictors and outcomes, enabling conclusions regarding which factors most strongly influence students' acceptance of this technology.

Independent samples t-tests were conducted to compare the mean scores of male and female participants across the four attitudinal dimensions: general acceptance, acceptance of facial expression recording, ethical concerns, and conditional acceptance. Under the assumptions of normal distribution and homogeneity of variance, the t-test was deemed suitable for comparing means of continuous outcome variables between two independent groups. The purpose of this analysis was to determine whether gender served as a significant predictor of attitudinal differences, or whether acceptance and concern were relatively gender-neutral. The results aimed to support or challenge hypotheses related to gender differences in technology perception, ethical reasoning, and risk sensitivity, thereby informing inclusive and demographically equitable practices in the implementation of educational technology.

To explore attitudinal differences across academic year levels, a one-way analysis of variance (ANOVA) was conducted among three groups, sophomores, juniors, and seniors. Under the assumptions of normality and homogeneity of variance, ANOVA was appropriate for comparing the means of continuous dependent variables across more than two independent groups. The analysis was applied to all four attitudinal dimensions to determine whether academic experience moderated students' acceptance of or concern about facial expression recognition technologies. Where statistically significant group differences were identified, post hoc multiple comparisons (Tamhane's T2) were conducted to determine which specific groups differed from one another. These comparisons provided insight into grade-level variations in attitudes and offered important evidence for guiding sample selection in future research.

3.2 Research Method of Study 2

3.2.1 An Overview of Study 2

This study used a within-participant comparative design in a tiered classroom at Shenyang Normal University. Students attended two course sessions in the same week, scheduled three days apart. In both sessions, they viewed the same time-based media art (TBMA) video, but the temporal structure differed (Table 3.3). At present, video documentation affords the most faithful representation of the essential features of TBMA works. In the first session, the video was screened continuously, and the students completed

Table 3.3: Procedures used for initial impression and delayed impression sessions (for study 2).

Stage	Initial Impression	Delayed Impression
Stage 1	10 min—Explained informed consent and experimental details; calibrated and tested recording equipment.	5 min—Explained experimental details; calibrated and tested recording equipment.
Stage 2	Four groups viewed different TBMA videos projected on screen; facial expressions were recorded.	Four groups viewed still frames extracted from previously shown videos; facial expressions were recorded.
Stage 3	30 s break.	2 min break.
Stage 4	20 min—Completed self-report questionnaire.	25 min—Completed self-report questionnaire.
After the experiment	5 min—Checked for any adverse reactions; gathered brief feedback.	10 min—Checked for any adverse reactions; explained creators' intent; collected participant feed-back.

an immediate written reflection. In the second session, the same work was presented as a guided reflective re-viewing that used curated still frames of fixed duration with brief prompts and short pauses to support recall and appraisal. This study was conducted to answer the following research question: How does guided reflective re-viewing of the same TBMA work after a three-day interval change students' emotional impressions?

Initial impression (II)—During the initial session, the participants directly viewed videos related to TBMA. Facial expressions of the participants were recorded using their private smart mobile devices (smartphones or tablets), while self-reported textual questionnaires were simultaneously administered. Delayed impression (DI)—As this study adopted an empirical research methodology and anticipated future extensions over a broader temporal scope, the delayed session was determined based on the actual academic calendar of Shenyang Normal University, where courses are held twice a week with a three-day interval. Accordingly, the delayed session was scheduled to take place on the third day following the conclusion of II. During the delayed session, the participants recalled the video content by viewing still frames extracted from the same footage. Their facial expressions were recorded using their private smart mobile devices, and they subsequently completed self-report textual questionnaires.

Given that the interval between the two viewing sessions in this study was limited to three days, the influence of forgetting was expected to be minimal. Consequently, the dominant emotional orientation formed during the initial impression was likely to persist, although its intensity could slightly decrease. In contrast, non-dominant emotions were expected to undergo more subtle modulation, becoming richer and more nuanced through guided reflective re-viewing. Based on this reasoning, two hypotheses were formulated to examine the consistency of and variation in emotions between the initial and delayed impressions:

Hypothesis 1. The dominant emotions in the II and DI conditions exhibit a high degree of consistency.

Hypothesis 2. The DI condition leads to a greater variety of emotional categories and more profound non-dominant emotions compared with II.

3.2.2 Participant Demographics and Ethical Considerations

In this study the number of participants was determined by the natural size of the intact class that participated in the course. The art appreciation course at Shenyang Normal University included twenty four third year painting majors, and all students in that class were invited and agreed to take part. This study therefore used a census of that specific learning group rather than selecting a small subset of a larger pool. This decision preserved the ecological validity of the teaching situation and reduced selection bias, because the emotional responses that were analysed reflected the full range of abilities and backgrounds present in the actual class (Table 3.4). In addition, having between twenty and thirty participants is often regarded as sufficient for the mean of many psychological variables to approximate a normal distribution (Altman and Bland, 1995). For these reasons, and in order to keep the natural structure of the class, the study retained the original group size of twenty four students instead of artificially increasing or decreasing the number of participants.

Table 3.4: Demographic information of participants.

Items	N	Percent (%)	Cumulative Percent (%)
Gender			
Male	5	20.8	20.8
Female	19	79.2	100.0
Duration of art learning			
3–5 years	14	58.3	58.3
5–10 years	9	37.5	95.8
More than 10 years	1	4.2	100.0

Moreover, according to the latest publicly available data released by the Chinese Ministry of Education in 2023 (http://en.moe.gov.cn), the proportion of female students in university art programs is relatively high. Therefore, the gender distribution of the sample in this study reflects the natural distribution, without any artificial balancing of gender proportions. Participants were randomly assigned into four groups of six students each ($M_{age} = 21.1$, $SD_{age} = 0.68$). Prior to their enrollment at Shenyang Normal University, they had under-gone systematic training in drawing, color theory, sketching, and art design.

Given that the experiment involved the recording of the participants' facial expressions, careful ethical considerations were implemented to protect the participants' privacy and rights. All participants signed a written informed consent form prior to participation. The participants were allowed to use their own smartphones or personal devices to ensure comfort and control during data collection. The recording duration was designed to be brief yet effective to minimize discomfort and intrusion. The participants were informed that the devices used had no unauthorized image enhancement or surveillance functions. They were also provided with opportunities to give feedback and express any concerns after the experiment. Data collection was conducted on-site, and all 24 participants provided the corresponding data. All participants had normal or corrected-to-normal vision.

3.2.3 Experimental Design and Procedure

This study involved four groups of participants, each of whom took part in two on-site experimental sessions conducted three days apart. Each group viewed a different video recording of a TBMA project sourced from YouTube. These videos varied in duration, stylistic characteristics, levels of abstraction, and visual complexity (Table 3.5).

Table 3.5: Information about the videos watched by each group.

Group	Video Information
G 1	Bill Viola, *Ascension*
G 2	Bill Viola, *The Raft*
G 3	Steve Reich, *Music for Pieces of Wood - Visualization*
G 4	Steve Reich, *Pendulum Music 1968*

The four TBMA works selected for this Study were chosen to represent two major modes of artistic experience. *Ascension* and *The Raft* by Bill Viola are works that rely on narrative progression and visually explicit human scenes. These pieces are known to elicit strong and immediate emotional reactions, making them appropriate for emotion recognition. In contrast, Steve Reich's *Music for Pieces of Wood* and *Pendulum Music* rely on rhythmic patterns, repetition, and gradual temporal development rather than story or imagery. These works encourage a different type of emotional engagement that often emerges through attentive listening, temporal anticipation, and reflective interpretation. Narrative imagery tends to evoke direct and easily identifiable emotions, while sound-focused or rhythm-focused pieces evoke more gradual, contemplative, or ambiguous emotional states (Meyer, 2024; Tan, 2013). TBMA education emphasizes sensory immersion, emotional engagement, and reflective reappraisal. Viola's works address emotional immediacy and sensory intensity, whereas Reich's works support concentrated attention and reflective processing. Using two works from each category also provided a balanced experimental design by reducing stylistic bias and allowing systematic comparison across different ways in which time-based artworks evoke emotions.

Both experimental sessions were conducted in the same tiered classroom (Figure 3.1).The videos were projected onto a 195 cm × 195 cm screen, maintaining the original resolution and aspect ratio as downloaded from YouTube. The setting of the tiered classroom and large screen ensured the controllability and effectiveness of the experiment to the greatest extent possible (Ravaja et al., 2004; Wang et al., 2009).

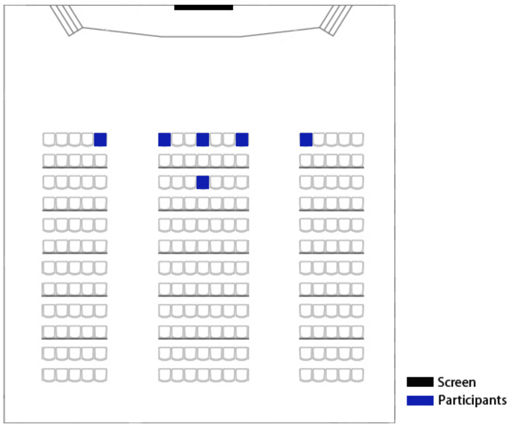

Figure 3.1: Top-down illustration of the experimental scene.

During the video viewing, each participant's facial expressions were recorded using their private smart mobile devices, and after watching the videos, they were asked to write down their thoughts and emotional responses regarding the content. During the recording of the participants' facial expressions, on-site researchers ensured the independence of each participant by excluding other participants from the footage. Similarly, the questionnaires were completed without any mutual influence between the participants.

3.2.3.1 Detailed Description of the Experimental Procedure (for Initial Impression)

The design of the experiment was based on the 45 min course time allocation at Shenyang Normal University.

Stage 1: The participants were introduced to the experimental procedures, including a five-minute informed consent briefing, followed by a five-minute explanation of the experimental process and assistance in setting up the facial expression recording devices.

Stage 2: Each group of six participants watched a different TBMA video while sitting in a tiered classroom facing the projection screen. Each participant's private smart mobile device was mounted on a tripod to record their facial expressions during the viewing.

Stage 3: To ensure that the participants retained a clearer memory for completing the questionnaire, only a 30 s break was given after the video presentation (BJORK, 2011; Semeniuta et al., 2016).

Stage 4: Based on emotion measurement under laboratory conditions, a questionnaire was designed (Wallbott and Scherer, 1989). The participants completed a self-report questionnaire on their emotional responses, writing a 300-word reflection on their experience while watching the video. The prompt for the self-report was as follows: "Please describe your feelings after watching this video (Describe the emotions you felt during it in as much detail as possible; in about 300 words)."

After the experiment, an open-ended group discussion was conducted in the tiered classroom to assess whether the participants experienced any discomfort or concerns. The participants were assured that they could discuss any personal concerns privately, ensuring their privacy and well-being.

3.2.3.2 Detailed Description of the Experimental Procedure (for Delayed Impression)

The second session followed a similar four-stage structure, with notable modifications in time allocation and procedural details.

In Stage 1, the informed consent explanation was omitted, as the participants had already provided consent in the first session. Instead, they only received a briefing on the specific procedures for the DI.

In Stage 2, the participants did not rewatch the full videos but instead viewed still frames extracted from the same TBMA videos at one-minute intervals. Each still frame was presented for 17 s, in accordance with the optimal viewing duration suggested by Brieber et al., allowing the participants to recall their previous viewing experience and associated emotions (Brieber et al., 2020).

In Stage 3, to stimulate the participants' recall and facilitate better cognitive processing while also preserving their short-term memory, the break duration was extended to two minutes (Cowan and AuBuchon, 2008).

In Stage 4, the participants were asked to complete the self-report emotional questionnaire, attempting to recall their initial impressions from the first session while writing a 300-word reflection on their experience. The prompt for the self-report was as follows: "Please describe your feelings after watching content related to this video? (Describe the emotions you felt during it in as much detail as possible; in about 300 words."

At the conclusion of the experiment, a detailed explanation of the artistic intent behind the TBMA works was provided, addressing the participants' questions and concerns regarding the artworks.

3.2.4 Data Preprocessing and Statistical Analysis of Facial Expression

Before conducting the data analysis, the facial expression data collected under the initial impression (II) and delayed impression (DI) conditions across all four groups were systematically preprocessed. The preprocessing steps were as follows:

Part 1: The participants' facial expression recordings were stored in folders categorized by group number and experimental condition (II vs. DI). Each participant's video file was labeled for easy identification.

Part 2: One frame per second was extracted from each participant's video using Python 3.10 (implemented in PyCharm 2023.3.3, an integrated development environment commonly used for machine learning applications) to capture close-up facial images. Each frame was timestamped and stored in sequential order. Given that macro-expressions typically last between 0.5 and 4 s and are readily observable, a sampling rate of 1 fps is generally sufficient for effective analysis, particularly when the original video is recorded at a standard frame rate (25–30 fps) (Ben et al., 2021). This approach was adopted in the present study to reduce the computational load while preserving essential emotional cues.

Part 3: A Convolutional Neural Network (CNN) emotion analysis model was applied to automatically classify the emotional states in each extracted image frame. The analysis was conducted in Python using PyCharm. The model had pre-trained CNN architecture based on PyTorch for facial expression recognition. This publicly available model was obtained from the GitHub website (Model link) and used to ensure reproducibility and benchmark-level performance. Each extracted frame was categorized into one of seven emotional states: anger, disgust, fear, happiness, sadness, surprise, and neutral (He et al., 2016). For each frame, the model generated probability scores for all emotion categories, and the dominant emotion per second was determined automatically based on the highest probability score. The entire classification process was completed by the algorithm without any human judgment or manual labeling. Finally, the frequency (in seconds) of each detected dominant emotion was recorded for every participant to support a quantitative comparison between experimental conditions.

Part 4: Since the video durations varied across the groups and the recording times for the II and DI conditions differed, a percentage-based normalization was applied to the emotion frequency data. This ensured unbiased comparisons across groups, eliminating potential distortions due to time differences (Aksu et al., 2019).

Part 5: After normalization, Min–Max normalization was applied to compute the dominant emotions for each participant under the II and DI conditions (Rafique et al., 2022). The dominant emotion of each participant was recorded for further statistical analysis.

To evaluate the consistency of the dominant emotions identified under the II and DI conditions across a three-day interval, Cohen's Kappa coefficient was applied as a statistical measure to evaluate the agreement between the II and DI dominant emotions (Höfel and Jacobsen, 2003). Since this study was based on quantitative analysis with a small sample size and slight variations between the groups, the normality of the data distribution was first assessed before analyzing the diversity and intensity of emotions under the II and DI conditions (Altman and Bland, 1995). After confirming that all data approximately followed a normal distribution, the homogeneity of variance across the four groups was further evaluated to ensure the validity of subsequent analyses. Given that the participants were divided into four groups, each viewing a different video, a one-way ANOVA was conducted to confirm the absence of significant baseline differences in emotional responses between the groups (Judd et al., 2017). After confirming that the results of the normality tests and the one-way ANOVA met the assumptions required for conducting a paired-sample t-test, to examine the emotional diversity, a paired-sample t-test was conducted to compare the range of emotions exhibited under the II and DI conditions. This test evaluates whether the mean difference between paired observations is statistically significant. Specifically, the analysis focused on whether the DI condition elicited a broader spectrum of emotional expressions compared with the II condition. Furthermore, to determine whether non-dominant facial emotions (NFE) were more prominent and intense under the DI condition compared with the II condition, the total proportion of non-dominant emotions observed for all participants was calculated for both conditions. A paired-sample t-test was then conducted to assess any significant differences (Blair and Higgins, 1985).

3.2.5 Data Preprocessing and Statistical Analysis of Textual Emotion

The textual emotion data collected under the II and DI conditions across all four groups were systematically preprocessed. The preprocessing steps were as follows:

Part 1: The participants' handwritten questionnaires were transcribed and stored in an electronic Word format. The files were categorized into folders by group number and experimental condition (II vs. DI), with each file labeled for identification purposes.

Part 2: Using PyCharm, emotion analysis was conducted based on a Chinese emotion lexicon, extracting emotion-related words from the participants' responses. The emotion analysis model classified words into seven emotional categories: happy, good, anger, sad, fear, dislike, and surprise

(Li et al., 2010). Additionally, the model identified the frequency of words associated with each emotion, as well as the stop words, total word count, and sentence count (Model link: cntext).

Part 3: Similar to the preprocessing of facial expression data, the dominant emotion in the textual responses was processed using Min–Max normalization. The dominant emotion for each participant was computed and recorded for the II and DI conditions.

These preprocessing steps ensured that the facial expression and textual emotion data were systematically analyzed while maintaining comparability across the experimental conditions.

This study analyzed self-reported emotional text data by assessing the consistency, diversity, and intensity of emotional expressions under the II and DI conditions. Cohen's Kappa coefficient was used to evaluate the consistency of the dominant emotions, ensuring the reliability of the categorized emotional data derived from the textual analysis. Similarly, after confirming that the results of the normality test and the one-way ANOVA satisfied the assumptions required for conducting a paired-sample t-test, emotional diversity was examined through a paired-sample t-test, which compared the range of emotions ex-pressed in self-reported textual responses across the two conditions. The goal was to determine whether the DI condition elicited a broader spectrum of emotional expression than the II condition. Differences in the intensity of non-dominant emotions were assessed using a paired-sample t-test on three key indicators: total non-dominant emotion words (TNEWs), representing the overall number of emotion-related words associated with non-dominant emotions; non-dominant emotion density (NED), reflecting the proportion of non-dominant emotion words within the total number of emotion words; and non-dominant emotion intensity per sentence (NEIPS), defined as the average number of non-dominant emotional words per sentence (Yadollahi et al., 2017).

3.3 Research Method of Study 3

3.3.1 An Overview of Study 3

The experimental design of Study 3 followed the same overall structure and emotion recognition procedures developed in Study 2. Each session was conducted twice per week, consisting of an Initial Impression (II) phase and a Delayed Impression (DI) phase held three days later (Table 3.6).

Table 3.6: Procedures used for initial impression and delayed impression sessions (for study 3).

Stage	Initial Impression	Delayed Impression
Stage 1	Participants viewed different TBMA videos projected on screen; facial expressions were recorded.	Participants viewed still frames extracted from previously shown videos; facial expressions were recorded.
Stage 2	30 s break.	2 min break.
Stage 3	25 min—Completed self-report questionnaire.	30 min—Completed self-report questionnaire.
After the experiment	5 min—Checked for any adverse reactions; gathered brief feedback.	5 min—Checked for any adverse reactions; explained creators' intent; collected participant feed-back.

The study was carried out over a four-week period as an extended case study to enhance the applicability of emotion recognition in TBMA education. Building on the findings of Study 2, Study 3 sought to further explore the temporal and affective dynamics of art appreciation within a more open classroom environment. Two research questions were conducted:

RQ1: Can Study 3 replicate the emotional change observed in Study 2?

RQ2: What is the relationship between the six emotional indicators and art appreciation within the classroom setting of Study 3?

3.3.2 Participants

Compared with Study 2, this case study involved a slightly larger group of third year undergraduate students enrolled in the College of Fine Art and Design at Shenyang Normal University Table 3.7. Although these participants were drawn from two classes, both classes belonged to the same major and participated in this case study as a merged group.

Initially, 32 students consented to participate in the study. However, due to absences and voluntary withdrawals during the course of the study, the final analytical sample comprised 29 participants, with a median age of 20.3 years (SD = 0.54). The composition of the participants also reflected gender representation trends commonly observed in Chinese higher art education.

Table 3.7: Demographic information of participants.

Items	N	Percent (%)	Cumulative Percent (%)
Gender			
Male	7	24.1	24.1
Female	22	75.9	100.0
Duration of art learning			
3–5 years	23	79.3	79.3
5–10 years	6	20.7	100.0

Therefore, consistent with the rationale adopted in Study 2, no artificial effort was made to balance the gender ratio, as doing so could have compromised the ecological validity of the case study. Prior to participation, all students were fully informed of the study's purpose, procedures, and data privacy protocols. Written informed consent was obtained from each participant. The research design and ethical procedures were reviewed and approved by the dean of Shenyang Normal University.

3.3.3 Experimental Design and Procedure

Study 3 built upon the methodological foundation of Study 2, extending both the experimental setting and analytical scope to enhance its applicability in media-art education. The study retained the same within-participant comparative design but introduced several contextual and procedural modifications. Whereas Study 2 involved four separate groups viewing different TBMA videos in a tiered classroom, Study 3 included twenty-nine participants drawn from two third-year art classes who jointly viewed the same video each week over four consecutive weeks. This change broadened the participant base and increased the diversity of artistic backgrounds within a unified classroom experience.

All four TBMA works were again selected from YouTube (Table 3.8), but the weekly videos differed from those used in Study 2 in order to provide new artistic appreciation. In contrast to the large tiered classroom with a 195 cm × 195 cm screen used previously, the present sessions were conducted in a smaller, open-studio environment equipped with a 150 cm × 150 cm screen. This arrangement created a more relaxed and interactive atmosphere that better reflected authentic studio-based learning in art education. *Egg Hatch* by Bill Viola presents a symbolic and contemplative visual experience. The slow unfolding of biological imagery allows viewers to enter a state of reflection and to experience emotions that emerge gradually through sustained attention. *Me and Beauty* by Studio Bilis Meurah focuses on identity, bodily presence, and self awareness. The work invites viewers to confront personal and social meanings related to beauty, appearance, and the experience of being observed. *Room Full of Mirrors* by CEBRA Architecture is centered on spatial perception and immersive visual environments. The reflective and fragmented space encourages exploration and often produces feelings associated with disorientation, curiosity, and perceptual uncertainty. The final work, *Digital Media Performance 2009* by Adrien M, highlights the interaction between the human body and digital imagery. The combination of movement, rhythm, and technological mediation creates an experience that engages attention through motion and through the shifting relationship between physical and digital forms. Each week was designed to present a clearly distinct type of artistic experience, including contemplative symbolic imagery, identity oriented performance, spatial and perceptual exploration, and technologically mediated motion. This variety allowed this study to observe how students responded emotionally when they encountered a different artistic emphasis each week.

Prior to the experimental sessions, participants had already been briefed about the project and had signed informed-consent forms. Therefore, the introductory explanation conducted in Stage 1 of Study 2 was omitted, allowing additional time for completing the expanded post-viewing questionnaire, which included new scales measuring understanding, interest, and liking toward each TBMA work.

The self-report questionnaire developed for Study 3 was designed to assess three dimensions (Brieber et al., 2020; Swami, 2013): Understanding, Interest, and Liking. Each dimension consisted of four statements rated on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). The Understanding scale measured participants' perceived cognitive grasp of the artwork ("I felt able to understand this artwork"), while the Interest scale captured their attentional and affective engagement ("This media artwork is fascinating"). The Liking scale reflected participants' affective preference and aesthetic enjoyment ("I find this artwork enjoyable"). In each dimension, three items were positively worded and one was reverse-coded ("This artwork was basically meaningless," "This media artwork is boring," "I dislike this artwork") to minimize response bias. Mean scores were computed for each dimension after reversing the negatively phrased items, yielding composite indices representing the participant's overall understanding, interest, and liking toward the viewed TBMA work.

Table 3.8: Information about the videos for each week.

Week	Video Information
W 1	Bill Viola, *Egg Hatch*
W 2	Studio Bilis Meurah, *Me and Beauty*
W 3	CEBRA Architecture, *Room Full of Mirrors*
W 4	Adrien M, *Digital Media Performance 2009*

3.3.4 Data Preprocessing and Statistical Analysis

The preprocessing was identical to that used in Study 2. All recorded videos were processed in Python 3.10 (PyCharm 2023.3.3) using the same convolutional neural-network (CNN) facial expressions model implemented in PyTorch. Each frame was analyzed at 1 fps and automatically classified into seven emotion categories, anger, disgust, fear, happiness, sadness, surprise, and neutral, according to the highest-probability label. The total duration of each dominant emotion was normalized to the total video length and further standardized through Min–Max normalization to enable comparison across participants and sessions.

Participants' written reflections were transcribed into electronic format and processed using the cntext Chinese emotion-lexicon model. Emotion-related words were identified and grouped into the same seven emotion categories as the facial-expression data. For each participant, the total frequency and proportion of emotion words were normalized using the same Min–Max procedure as in Study 2.

Statistical analysis for RQ1:

To examine whether the emotional change patterns identified in Study 2 could be replicated in Study 3, a series of statistical analyses were conducted following the same analytical framework established in the previous study. The analysis first used Cohen's Kappa to examine the consistency of dominant emotions between the II and DI sessions. Subsequently, for each week, participants' emotional data was compared using paired-sample t-tests. When the assumption of normality was not met, the corresponding Wilcoxon signed-rank test was used. These tests assessed whether participants' emotional intensity significantly increased, decreased, or remained consistent after the three-day reflective interval.

By applying the same procedures and normalization methods as in Study 2, the analysis ensured full methodological consistency and allowed direct comparison of temporal emotional patterns across studies. The results of these analyses provide empirical evidence on whether the emotional modulation observed in Study 2, particularly the amplification of non-dominant emotions during delayed reflection, was reproducible in the expanded classroom context of Study 3.

Statistical analysis for RQ2:

Difference-score regression analyses were performed in SPSS to directly assess the relationship between changes in emotion and changes in appreciation (Laird and Weems, 2011; Shek and Ma, 2011). For each participant and week, the differences between II and DI were calculated as Emotion = $\mathrm{DI}-\mathrm{II}$ for each of the six emotion indicators and Outcome = $\mathrm{DI} - \mathrm{II}$ for Understanding, Interest, and Liking. Each regression model used Emotion as the predictor and Outcome as the dependent variable, thereby testing whether increases in emotional intensity were associated with corresponding gains in appreciation outcomes. This approach provided an intuitive validation of the temporal association between emotional change and the enhancement of students' understanding, interest, and liking of time-based media artworks.

To further investigate the relationship between emotions and art appreciation, linear mixed models (LMM) were conducted in SPSS to examine the relationships between emotional indicators and the outcomes of understanding, interest, and liking. LMM were selected instead of traditional analyses of variance because they allow the simultaneous modeling of both categorical and continuous predictors while accounting for within-subject repeated measures (Baayen et al., 2008; Judd et al., 2012). Study 3 placed greater emphasis on the overall differences between sessions (II and DI) as well as on individual variations across participants. Moreover, as the mean values of artworks do not permit generalization of results to other stimuli drawn from the same population of artworks (Westfall et al., 2014), it was essential to account for both participant- and stimulus-level variability, particularly in studies involving artworks. By modeling Time (II vs. DI), emotional indicators, and their interaction (Time × Emotional Indicator) as fixed effects, the analysis not only assessed whether emotions predicted students' appreciation but also whether the strength of these relationships varied over time. The participant variable was specified as a random intercept to capture intra-individual dependencies, and a compound-symmetric covariance structure was applied to model correlated residuals between the two sessions.

The LMM modeling procedure proceeded as follows. First, a baseline model including only random intercepts for participants was tested. Subsequently, fixed effects of Time, emotional indicators, and their interaction (Time × Emotional Indicator) were constructed. Each of the six standardized emotional indicators was analyzed in a separate model to avoid multicollinearity among six emotional predictors. Prior to modeling, all emotional indicators were grand-mean centered within each model to minimize multicollinearity between Time and the emotional predictor itself, and to facilitate interpretation of the main effects at the mean emotional level. Likelihood ratio tests were used to determine whether adding the interaction term improved model fit. The significance of fixed effects was evaluated using Type III tests of fixed effects, and the interpretability of significant interactions was further examined through post-hoc slope analyses at each time level (II and DI) (Barr et al., 2013). This modeling strategy allowed for the assessment of whether emotional intensity was consistent across the two viewing sessions or whether the emotional influence was time-dependent, reflecting the dynamic changes of emotional engagement and cognitive reappraisal during repeated exposure to TBMA works.

Chapter 4

Results of Feasibility of Emotional Recognition and Emotional Change Trends(Study 1 and Study 2)

4.1 Results of Study 1

4.1.1 Reliability and Validity of Instrument

To evaluate the internal consistency of the questionnaire developed in this study regarding students' perceptions of smartphone-based facial expression recording, a reliability analysis was conducted using Cronbach's alpha. As shown in Table 4.1, the overall Cronbach's alpha coefficient for the scale was .797, indicating a good level of internal consistency (Nunnally, 1994), particularly appropriate for instruments in the early stages of development. The questionnaire consisted of 12 items, each rated on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Table 4.1: The overall reliability of the instrument.

	Cronbach's alpha	Number of items
Overall scale	.797	12

The relatively high alpha coefficient supports items that consistently measure students' attitudes, expectations, and concerns about the use of smartphone-based facial expression recording in classroom settings.

As part of the scale refinement process, an item-level reliability analysis was conducted to examine each item's contribution to the overall internal consistency. Table 4.2 presents the Cronbach's alpha values if individual items were deleted. Most values ranged between .745 and .811, indicating that no single item substantially reduced the scale's reliability.

Table 4.2: Item analysis based on Cronbach's Alpha if item deleted.

Question No.	Cronbach's Alpha if question deleted
Q5	.794
Q6	.791
Q7	.792
Q8	.773
Q9	.745
Q10	.753
Q11	.774
Q12	.777
Q13	.762
Q14	.798
Q15	.808
Q17	.811

Notably, the alpha values associated with Q15 (.808) and Q17 (.811) were slightly higher than the overall scale value, suggesting that the removal of either item might slightly enhance internal consistency. However, this potential increase was not substantial enough to warrant item removal, especially given the exploratory nature of this newly developed instrument. Conversely, the removal of items such as Q9 (.745) and Q10 (.753) resulted in lower alpha values, implying that these items made a positive contribution to the scale's consistency. Given the exploratory aim of the study, to construct a simple and accessible tool for assessing students' initial perceptions of mobile-based facial expression recording in educational contexts, the results are encouraging. The overall reliability coefficient (.797), together with the item-level alpha values, indicates that the scale demonstrates acceptable psychometric properties for use in early-stage research.

To evaluate the construct validity of the questionnaire, an Exploratory Factor Analysis (EFA) was conducted following the overall reliability analysis. The EFA confirmed a multidimensional structure and provided insights into the internal coherence and theoretical foundation of the instrument.

As presented in Table 4.3, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was .795, indicating a satisfactory level of inter-item correlations for factor analysis (Shrestha, 2021). Bartlett's Test of Sphericity was statistically significant ($\chi^2(66) = 890.004$, $p < .001$), confirming that the correlation matrix is appropriate for dimension reduction. Together, these results support the suitability of the dataset for EFA (Yong et al., 2013).

Table 4.3: KMO and Bartlett's Test.

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.795
Bartlett's Test of Sphericity	Approx. Chi-Square	890.004
	df	66
	Sig.	.000

EFA, combined with the extraction and interpretation of principal components, identified a three-component structure, reflecting distinct underlying dimensions of student perceptions. The component matrix (Table 4.4) presents the factor loadings that support the interpretation of each component.

In general, factor loadings greater than 0.4 or 0.5 are considered meaningful contributions to a given factor, while loadings below this threshold typically indicate weaker item-factor associations (Williams et al., 2010). For clarity, this study applied a cutoff value of 0.5 during the analysis; loadings below this threshold were not displayed in the component matrix.

Table 4.4: Rotated component matrix.

	Component		
Question No.	1	2	3
Q5	.808		
Q6	.805		
Q7	.665		
Q8		.507	
Q9		.924	
Q10		.884	
Q11		.821	
Q12		.587	
Q13		.783	
Q14			.694
Q15			.694
Q17			

Component 1: This component includes high loadings for Q5 (.808), Q6 (.805) and Q7 (.665). These items relate to general approval, perceived usefulness, and support for mobile-based video/photo recording in educational projects. Collectively, this dimension represents students' pragmatic acceptance of the technology.

Component 2: The second factor comprises Q8 (.507), Q9 (.924), Q10 (.884), Q11 (.821), Q12 (.587) and Q13 (.783). These items concern students' emotional reactions, learning-related awareness, and psychological comfort when being recorded. This dimension captures affective engagement and reflective awareness, central to understanding how emotion recognition technology impacts learners cognitively and emotionally.

Component 3: The third component includes Q14 (.694) and Q15 (.694), which both focus on ethical risks and fairness issues related to analyzing facial expressions. These loadings suggest a coherent dimension representing students' critical concerns about the implications of using facial analysis in educational assessment contexts.

It is important to emphasize that Q17 represents a conditional statement within the instrument, directly dependent on the multiple-choice preferences outlined in Q16. Unlike Q15, which is framed as another fairness-related item, Q17 offers a succinct judgment of conditional acceptance, reflecting a moderate and context-sensitive stance. Although the semantic content of Q17 overlaps with the thematic focus of Q14 and Q15, it did not load significantly on any of the three extracted components, and reliability analysis indicated that its inclusion slightly reduced the overall Cronbach's alpha. This statistical finding supports the conceptual and structural distinctiveness of Q17. From a sequential perspective, Q17 follows a non-Likert, multiple-choice item (Q16), positioning it as a reflective follow-up rather than a standard component of the main scale. Functionally, Q17 assesses ultimate acceptance, contingent upon whether the student's preferences stated in Q16 are respected. Accordingly, the final dimensional structure of the scale is represented as follows:

General level of acceptance: Q5, Q6 and Q7.

Acceptance of facial expression recording: Q8, Q9, Q10, Q11, Q12 and Q13.

Ethical and fairness concerns: Q14 and Q15.

Conditional level of acceptance (independent): Q17 (derived from and dependent on Q16).

Collectively, these items extend the functionality of the scale by providing design-related feedback and capturing thresholds of acceptance regarding the ethical and personalized use of technology in educational contexts.

4.1.2 Results of Device Ownership and General Attitudes toward Educational Use

This study initially recruited 202 undergraduate students on a voluntary basis to investigate their perceptions of using smart mobile devices to capture facial expressions in educational settings. However, data from three participants were incomplete and could not be reliably identified, resulting in their exclusion from the final analysis. Therefore, the descriptive statistics reported here are based on a total of 199 valid responses. All participants were full-time undergraduate students enrolled at a university and fell within the expected age range for traditional college students. Demographic information collected included age, gender, and academic standing(see Table 4.5).

Table 4.5: Descriptive statistics of participants' ownership of smart mobile devices.

		N	Percent(%)	Cumulative Percent(%)
Gender				
	Male	57	28.6	28.6
	Female	142	71.4	100.0
Age				
	19	20	10.1	10.1
	20	57	28.6	38.7
	21	76	38.2	76.9
	22	26	13.1	89.9
	23	14	7.0	97.0
	24	6	3.0	100.0
Grade				
	2nd year (Sophomore)	71	35.7	35.7
	3rd year (Junior)	92	46.2	81.9
	4th year (Senior)	36	18.1	100.0
Smart mobile device ownership				
	Yes	**199**	**100.0**	**100.0**
	No	0.0	0.0	0.0

The sample consisted of 57 male participants (28.6%) and 142 female participants (71.4%), indicating a gender imbalance favoring female representation. This imbalance reflects the distribution of students within relevant academic disciplines, where female students are typically overrepresented in the arts. In terms of age, participants ranged from 19 to 24 years old. The largest proportion of students were aged 21 (n = 76, 38.2%), followed by those aged 20 (n = 57, 28.6%) and 22 (n = 26, 13.1%). The remaining participants were aged 19 (n = 20, 10.1%), 23 (n = 14, 7.0%), and 24 (n = 6, 3.0%). The cumulative percentages indicated that by age 22, nearly 90% of the sample was accounted for, suggesting a relatively homogeneous age distribution typical of lower-division undergraduate populations. Age serves not only as a demographic descriptor but also as a conceptually relevant factor in explaining technology use behaviors. Younger individuals, particularly those within this age range, tend to exhibit greater digital fluency and higher levels of acceptance of mobile technologies in educational contexts. Regarding academic standing, participants were drawn from three distinct year groups. The second-year students comprised 35.7% of the sample (n = 71), while the third-year students formed the largest group, accounting for 46.2% (n = 92). Fourth-year students (senior undergraduates) made up 18.1% of the sample (n = 36). This grade-level distribution closely mirrored the age distribution and reinforced the notion that the majority of participants were situated in the mid-phase of their undergraduate education. The concentration of second- and third-year students (over 80% of the sample) suggests that these individuals had accumulated sufficient classroom experience to have formed informed opinions on the use of educational technologies, including smart mobile devices. Moreover, their academic maturity may have contributed to more reflective responses regarding data privacy, ethical use, and the potential educational benefits associated with mobile-based facial expression recording.

Based on the demographic information, the descriptive analysis revealed that all 199 valid participants reported owning at least one smart mobile device, resulting in a device ownership rate of 100.0%. This finding provides compelling evidence for the widespread use of smartphones or tablets among undergraduates within the sampled institution. The complete saturation of mobile device ownership corroborates previous literature suggesting that mobile technology has become an essential and ubiquitous tool in students' academic and personal lives (Sung et al., 2016). As every participant had access to at least one smart mobile device, the responses could be interpreted without concern for access-related bias or digital exclusion. This also allowed for the meaningful administration of subsequent questionnaire items that assumed familiarity with the functionality and usage of such technologies, such as those addressing comfort, willingness, and perceived fairness. Given that the study focused specifically on using students' personal devices, rather than institutional equipment or controlled laboratory setups, to record facial expressions during educational activities, the 100% ownership rate was particularly significant. Because all respondents indicated access to mobile technology, it is evident that technical feasibility is not a limiting factor in implementing mobile-based emotion recognition systems in educational environments. Instead, the focus shifts to other variables such as emotional readiness, ethical comfort, informed consent, and pedagogical alignment.

The scores for each dimension were calculated by aggregating responses to the corresponding items identified in the validated factor structure. As presented in Table 4.6 and further visualized in the accompanying bar chart(see Figure 4.1), the results reveal substantial variation in the attitudes of the 199 valid participants.

Table 4.6: Participants' general attitudes.

	N	Minimum	Maximum	Mean	Std. Deviation
General level of acceptance	199	3.00	5.00	**4.50**	0.62
Acceptance of facial expression recording	199	1.83	4.67	**2.87**	0.61
Ethical and fairness concerns	199	1.50	5.00	**4.30**	0.69
Conditional level of acceptance (independent)	199	4.00	5.00	**4.50**	0.50

Beginning with general level of acceptance, participants expressed a strong positive attitude toward the use of smart mobile devices for academic purposes, with a mean score of 4.50 (SD = 0.62), and scores ranging from 3.00 to 5.00. This high mean indicates that students broadly perceive mobile devices as beneficial and acceptable learning tools. Items corresponding to this dimension (Q5 to Q7) reflect support for conventional academic uses, perceived educational benefits, and the use of such devices for classroom-based media capture (photographing course materials or recording videos). The high scores in this dimension suggest that mobile devices are deeply integrated into students' learning routines and are not regarded as disruptive or inappropriate. On the contrary, they are viewed as valuable educational tools, likely due to their flexibility, portability, and alignment with contemporary learning practices such as blended learning, on-demand access to information, and interactive engagement. The relatively low standard deviation (SD = 0.62) further indicates broad consensus across participants, regardless of age, gender, or academic year, suggesting that a normative level of technological acceptance has been reached among undergraduate students (Years 2 through 4). In contrast, participants displayed more

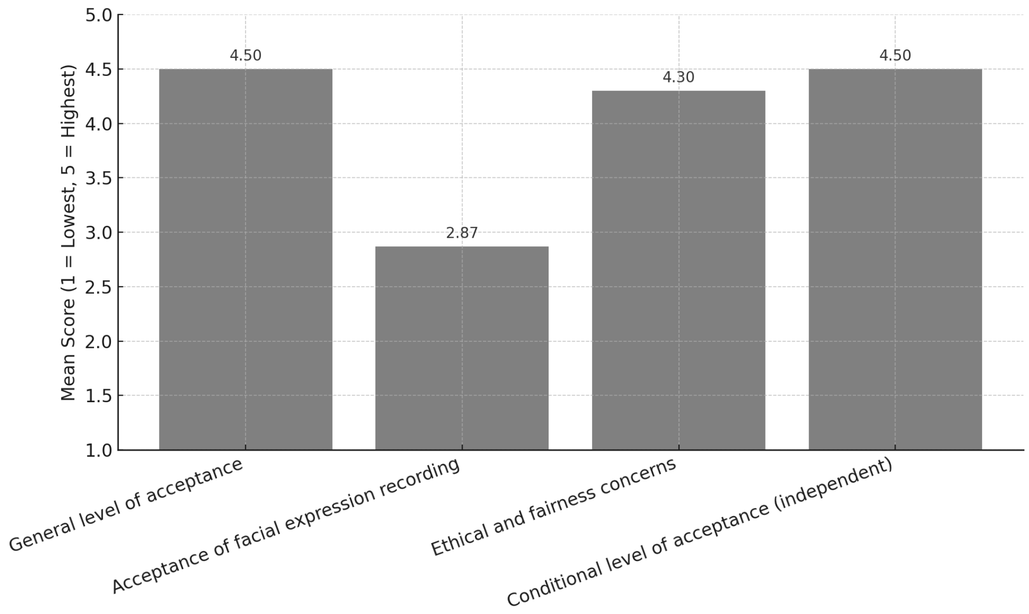

Figure 4.1: The mean score of participants' attitude

ambivalent attitudes toward the specific use of mobile devices for recording facial expressions, with a mean score of only 2.87 (SD = 0.61), ranging from 1.83 to 4.67. This was the lowest mean among the four dimensions, suggesting that while students generally support the educational use of smart devices, they feel significantly less comfortable with their application in emotionally intrusive or privacy-sensitive contexts. This dimension, based on items Q8 through Q13, includes beliefs regarding whether facial expression recording enhances learning, accurately reflects emotional states, affects concentration, or raises concerns about discomfort and misinterpretation. The standard deviation (SD = 0.61), although modest, indicates a considerable degree of dispersion in responses, reflecting diverse views potentially influenced by individual experiences, privacy orientations, and familiarity with emotion recognition technologies.

The stark contrast between general acceptance and the acceptance of facial expression recording highlights a critical boundary: although students may value mobile technologies as learning aids, they remain skeptical or uneasy about their role in monitoring or analyzing emotional states. This hesitation may stem from concerns over the intrusiveness of real-time monitoring, fear of misjudgment, or discomfort with the novelty of such practices in educational settings. Thus, the findings reveal a key attitudinal fault line between technological enthusiasm and ethical hesitation. Supporting this interpretation is the third attitudinal dimension, ethical and fairness concerns, which yielded a relatively high mean score of 4.30. On average, this indicates that participants are highly concerned about the ethical implications of using facial expression analysis in the classroom, including issues related to bias, surveillance, consent, and equitable treatment. The high mean scores for items Q14 and Q15 demonstrate that such concerns are not limited to a vocal minority but are widely shared across the sample. Additionally, the slightly elevated standard deviation (SD = 0.62) compared to the first two dimensions suggests a relatively consistent distribution of ethical concerns among participants. The coexistence of high general acceptance of devices (M = 4.50) and strong ethical hesitation (M = 4.30) paints a complex portrait of student attitudes, one in which recognition of technological functionality does not equate to endorsement of all its potential applications.

The mean score for participants' Conditional level of acceptance (independent), measured independently through Q17, was 4.50 (SD = 0.50), identical to their mean score for general acceptance of using smart mobile devices for educational purposes (M = 4.50, SD = 0.62). This finding suggests that, when participants' preferred ethical and procedural conditions, outlined in Q16, are hypothetically satisfied, their support for facial expression recording aligns with their high level of comfort with mobile technology in classroom contexts. This is particularly noteworthy given that participants' initial, unconditional acceptance of facial expression recording was considerably lower, with a mean of only 2.87 (SD = 0.61). The pronounced discrepancy highlights significant discomfort or resistance toward being recorded under default conditions, underscoring the transformative impact of ethical design on student acceptance. Furthermore, participants expressed strong concerns regarding ethics and fairness (M = 4.30, SD = 0.687), further clarifying the low level of unconditional acceptance. However, the fact that conditional acceptance rebounded to 4.50, despite these concerns, indicates that such apprehensions are not immutable. Rather, they can be effectively addressed through thoughtful design and transparent communication. Responses to Q16 reinforce this interpretation (see Figure 4.2). The most frequently endorsed conditions included: the provision of clear explanations and voluntary opt-out (n = 191), the ability for students to use their own devices (n = 172), and the availability of beautification features (n = 161). These preferences highlight students' emphasis on autonomy, personalization, and clarity, factors that, when respected, lead to high levels of conditional acceptance.

While students demonstrated a high degree of general acceptance toward smart mobile devices, they remained cautious about emotionally sensitive applications unless certain conditions were met. The stark contrast between

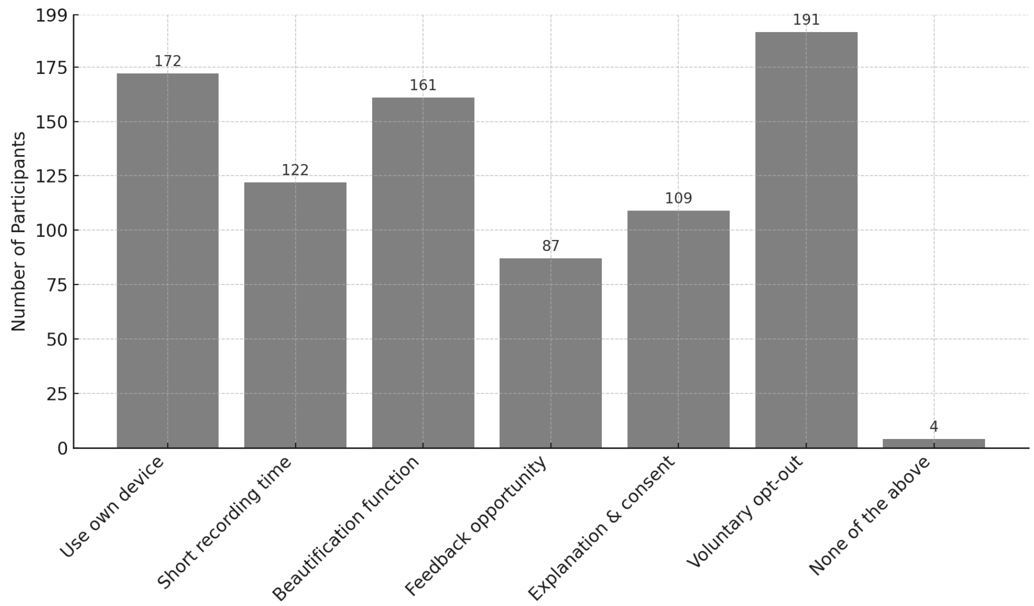

Figure 4.2: Participants' answers to Q16

unconditional and conditional acceptance scores indicates that students are not inherently opposed to facial expression recording; rather, they seek trustworthy and student-centered implementation protocols. This insight offers clear direction for subsequent Studies 2 and 3: in order to harness the pedagogical potential of facial expression recognition technologies, ethical design features must not be treated as secondary considerations. Instead, they should be regarded as primary determinants of student willingness and project feasibility.

4.1.3 Results on Predictors of Usage Intention and Demographic Differences

Pearson product-moment correlation analysis was conducted to address RQ2. As presented in Table 4.7, the results offer insights into the strength and direction of associations among the three validated dimensions previously identified through factor analysis: (1) general level of acceptance, (2) acceptance of facial expression recording, and (3) ethical and fairness concerns. These relationships provide further support for the patterns observed in the earlier descriptive findings and contribute to a more nuanced understanding of students' attitudes toward the implementation of affective technologies in educational contexts. All three variables were treated as continuous measures derived from Likert-type composite scales and met the assumptions for Pearson correlation, including linearity, interval-level data, and the absence of extreme outliers.

Table 4.7: Correlation matrix between general acceptance, facial expression recording acceptance and ethical/fairness concerns.

	General level of acceptance	Acceptance of facial expression recording	Ethical and fairness concerns
General level of acceptance	1		
Acceptance of facial expression recording	0.236**	1	
Ethical and fairness concerns	0.136	0.301**	1

** $p < .01$.

A statistically significant positive correlation was found between general acceptance and acceptance of facial expression recording, with a Pearson correlation coefficient of 0.236 ($p < .01$). This association suggests a consistent and meaningful trend (Cohen et al., 2013): students who generally support the use of smart mobile devices in educational contexts are more likely to accept the idea of using these devices to record facial expressions. Importantly, while these two constructs are distinct, supported by their separation in factor analysis, they are not independent. This implies that students' positive experiences with or habitual use of mobile devices for learning (accessing educational apps or recording lectures) may form a conceptual foundation that facilitates openness to more advanced applications of the technology, including affective monitoring. However, the modest strength of this correlation also reinforces the notion that general technological acceptance does not automatically translate into acceptance of emotionally or ethically sensitive applications of the same tools. This gap aligns with the descriptive results, which showed that although the average score for general acceptance was relatively high (M = 4.50), the mean score for acceptance of facial expression recording was substantially lower (M = 2.87), suggesting that intervening variables, such as privacy concerns or discomfort with surveillance, moderate the relationship between general openness and specific consent. The correlation between acceptance of facial expression recording and ethical and fairness concerns further substantiates this interpretation, yielding a statistically significant positive correlation of 0.301 ($p < .01$). At first glance, this result may appear counterintuitive, as greater ethical concern might be expected to correlate with lower acceptance of recording practices. However, the positive direction of the correlation can be interpreted as indicating that students who are more attuned to ethical issues are also more capable of forming nuanced, conditional forms of acceptance. In other words, heightened ethical awareness may not lead to outright rejection but instead foster a more critical and engaged stance toward implementation. These students may support facial expression recording provided that ethical safeguards are clearly articulated, as evidenced by the high conditional acceptance score (M = 4.50). This interpretation is consistent with prior analysis of responses to Q16, where the most frequently endorsed conditions for acceptance included informed consent, the use of personal devices, and the availability of beautification features, each addressing distinct ethical dimensions (i.e., autonomy, control, and self-presentation, respectively). Thus, the positive correlation should not be understood as uncritical endorsement, but rather as evidence of a standards-based, reflective approach to technology adoption. This finding aligns with research in technology ethics (Van den Hoven et al., 2012), which suggests that awareness of ethical issues can stimulate constructive dialogue and design-oriented thinking, rather than passive avoidance or resistance. In contrast, the correlation between general acceptance and ethical and fairness concerns was positive but not statistically significant, with a Pearson correlation coefficient of 0.136. This result indicates that broad student support for mobile devices in education is largely independent of concerns about fairness or ethical implications related to facial expression recording. Although both constructs were associated with relatively high mean scores (general acceptance M = 4.50; ethical/fairness concerns M = 4.30), the lack of a significant statistical relationship suggests that students can simultaneously hold strong positive views about educational technologies while maintaining ethical reservations about specific applications, without perceiving these positions as contradictory. This disconnect may be attributable to the context-specific nature of ethical concerns, rather than general aversion to technology. For instance, students may enthusiastically use mobile devices for collaborative note-taking or accessing digital resources, yet still feel uncomfortable with those same devices being used to infer emotional states without sufficient transparency or control.

Overall, the correlation patterns reported in Table 4.7 reveal a multi-layered structure of relationships that reflects the attitudinal complexity recorded in the descriptive statistics. A statistically significant but moderate correlation exists between general acceptance and acceptance of facial expression recording, while a stronger relationship was found between acceptance of facial expression recording and ethical concerns. In sum, the correlation analysis for RQ1 supports and extends the descriptive findings by providing empirical evidence of both interdependence and differentiation among general technological acceptance, specific acceptance of facial expression recording, and ethical/fairness considerations. While students who embrace mobile learning are generally more likely to support facial expression recording, this support is conditional, reflective, and closely tied to the ethical framing of technological use.

The results, summarized in Table 4.8, revealed a statistically significant regression model, with an ANOVA F-value of 14.537 ($p < .001$) and an R^2 value of 0.129. This indicates that approximately 12.9% of the variance in students' acceptance of facial expression recording can be explained by the two predictors included in the model: general acceptance and ethical and fairness concerns. Although the R^2 value may appear modest, it is consistent with norms in social science research, where explained variances between 10% and 20% are typically considered meaningful for models predicting complex psychological or behavioral outcomes (Aiken et al., 1991; Hair et al., 2019). The statistical significance of the overall model justifies further interpretation of the individual coefficients.

Table 4.8: Linear regression predicting acceptance of facial expression recording.

	Unstandardized Coefficients		Standardized Coefficients			
	B	Std. Error	Beta	t	Sig.	VIF
(Constant)	0.942	0.369		2.554	.011	
General level of acceptance	0.196	0.066	0.199	2.957	.003	1.019
Ethical and fairness concerns	0.243	0.060	0.274	4.067	.001	1.019
Model - R^2					0.129.	
ANOVA - F					14.537	
ANOVA - p					$< .001$	

Dependent Variable: Acceptance of facial expression recording

General acceptance of smart mobile devices yielded a positive unstandardized regression coefficient (B = 0.196, SE = 0.066, p = .003) and a standardized beta coefficient of 0.199. This suggests that for each one-unit increase in general acceptance, the predicted score for acceptance of facial expression recording increases by 0.196 units, holding other variables constant. This finding aligns well with the earlier Pearson correlation analysis, which showed a significant but moderate positive relationship between general acceptance and facial expression recording acceptance (r = 0.236, $p < .01$). The regression coefficient builds upon the earlier descriptive statistics, where general acceptance was found to have a high mean score (M = 4.50), suggesting that students who frequently use mobile technology in educational settings may be more inclined to extend this acceptance, at least partially, to more novel and potentially intrusive applications such as emotional data collection. However, the relatively small standardized beta coefficient indicates that general technological endorsement is not the primary driver of students' acceptance of emotion-related data collection. Instead, it serves as a moderate but meaningful baseline disposition, which may influence attitudes when other variables, such as ethical comfort and personal control, are favorable. More compellingly, ethical and fairness concerns exhibited a stronger and statistically significant relationship with the outcome variable, with an unstandardized coefficient of B = 0.243 (SE = 0.060, $p < .001$) and the highest standardized beta coefficient in the model at 0.274. This finding suggests that, assuming such concerns are acknowledged and addressed, students reporting higher levels of ethical and fairness considerations are also more likely to express acceptance of facial expression recording. At first glance, this may appear counterintuitive, as greater ethical concern might be expected to correspond with lower acceptance. However, this interpretation is clarified by earlier descriptive and conditional acceptance analyses. Specifically, as shown in Table 4.6, unconditional acceptance of facial expression recording was relatively low (M = 2.87), whereas conditional acceptance (measured in Q17), assessed after participants selected their preferred ethical safeguards in Q16, rose to a level equivalent to general acceptance (M = 4.50). This dramatic shift underscores that students' ethical concerns do not necessarily manifest as opposition to technology itself, but rather as demands for responsible, transparent, and consent-driven implementation. Thus, students with strong ethical concerns do not outright reject the use of emotion-sensing tools; instead, they are more attuned to the conditions under which such tools may be acceptable. As previously reported in the correlation analysis, the strong positive association between ethical concerns and acceptance of facial expression recording (r = 0.301, $p < .01$) further reinforces this interpretation.

The importance of these findings is further emphasized by the low multicollinearity within the model, as indicated by the variance inflation factor (VIF) values of 1.019 for both predictors, well below the commonly accepted threshold of 10, or even the more conservative threshold of 2.5. This confirms that the two predictors contributed independently to the model and did not redundantly measure the same construct (Hair et al., 2019). They reflect two qualitatively distinct sources of influence on student attitudes: one grounded in instrumental and experiential familiarity with educational technologies, and the other rooted in ethical reasoning and value-based judgment. The statistical independence of these dimensions is consistent with earlier theoretical and empirical insights, particularly the weak and non-significant correlation between general acceptance and ethical concerns ($r = 0.136$), suggesting that students can simultaneously embrace technology while critically examining its implications, without perceiving these positions as contradictory. This multidimensional cognitive stance is especially important for educational researchers and instructional designers, as it indicates that students' resistance to facial expression recording is not a rejection of innovation per se, but a call for a more thoughtful integration of technological capability with ethical transparency and user autonomy. The regression results strongly support the argument that ethical design is not merely a normative supplement to educational technology implementation but a core determinant of acceptance, potentially even outweighing general technological enthusiasm. While broad acceptance of mobile devices does contribute to students' openness to emotion-sensing tools, the greater influence of ethical and fairness considerations implies that any initiative involving facial expression analysis must prioritize informed consent, data privacy, clear purpose articulation, and the option to opt out. This conclusion is once again aligned with participants' responses to Q16, where the most commonly endorsed conditions for acceptance included "clear explanation and informed consent" ($n = 191$), "use of personal devices" ($n = 172$), and "availability of beautification features" ($n = 161$). These preferences reflect a widespread desire for control, clarity, and emotional safety, factors that collectively underpin the high level of conditional acceptance observed in Q17 ($M = 4.50$).

To further address RQ2, independent samples t-tests were conducted on the four validated attitudinal dimensions. As shown in Table 4.9, the results revealed no statistically significant gender differences across any of the measured dimensions. The sample consisted of 57 male and 142 female participants, and the analysis compared their mean responses regarding.

First, with respect to general acceptance of smart mobile devices in educational contexts, male participants reported a mean score of 4.49 ($SD = 0.60$), while female participants reported a slightly higher mean of 4.50 ($SD = 0.63$). This difference was negligible and not statistically significant, $t(197) = -0.090$, $p = .928$, indicating that male and female students were equally likely to support the academic use of mobile devices. Similarly, no significant gender difference was found in the dimension of facial expression recording ac-

Table 4.9: Independent samples t-test for gender differences.

	Male	Female	t_{197}	p
General level of acceptance	4.49 ± 0.60	4.50 ± 0.63	-0.090	.928
Acceptance of facial expression recording	2.93 ± 0.60	2.85 ± 0.62	0.853	.395
Ethical and fairness concerns	4.22 ± 0.76	4.34 ± 0.66	-1.039	.302
Conditional level of acceptance (independent)	4.49 ± 0.50	4.50 ± 0.50	-0.111	.911

ceptance, which had the lowest overall mean score in the descriptive analysis. Male students reported a mean of 2.93 (SD = 0.60), while female students had a slightly lower mean of 2.85 (SD = 0.62). The t-test result, t(197) = 0.853, p = .395, confirmed that this minor difference was not statistically significant. These findings suggest that gender does not systematically influence students' comfort or discomfort with being recorded for emotional analysis, and both groups exhibit comparably cautious attitudes toward this technology. In the domain of ethical and fairness concerns, female students reported a slightly higher level of concern (M = 4.34, SD = 0.66) than their male counterparts (M = 4.22, SD = 0.76); however, this difference also failed to reach statistical significance, t(197) = -1.039, p = .302. This result indicates that ethical awareness surrounding facial expression analysis in education is broadly shared across genders. Finally, the t-test for conditional acceptance also showed no gender-based difference. The mean scores for both male and female participants were nearly identical (M = 4.49 and M = 4.50, respectively; both with SD = 0.50), with a non-significant result, t(197) = -0.111, p = .911. This further reinforces the notion that, when ethical safeguards such as voluntary participation, purpose clarification, and the use of personal devices are in place, students of all genders are equally likely to express high levels of acceptance. In other words, regardless of demographic differences, ethically designed implementation serves as a universally effective pathway to acceptance.

A one-way analysis of variance (ANOVA) was conducted on the four validated attitudinal dimensions. As summarized in Table 4.10, beginning with general acceptance, the ANOVA results indicated a statistically significant difference among academic years, F = 10.937, $p < .001$. Sophomore students reported the lowest mean score (M = 4.25, SD = 0.67), followed by seniors

Table 4.10: One-way ANOVA differences across grades.

	Sophomore	Junior	Senior	F	p
General level of acceptance	4.25 ± 0.67	4.70 ± 0.51	4.47 ± 0.61	10.937	.001
Acceptance of facial expression recording	2.64 ± 0.50	3.00 ± 0.64	3.01 ± 0.61	9.829	.001
Ethical and fairness concerns	4.27 ± 0.72	4.36 ± 0.64	4.22 ± 0.75	0.706	.495
Conditional level of acceptance (independent)	4.45 ± 0.50	4.55 ± 0.50	4.54 ± 0.50	1.104	.334

Multiple Comparisons.
General level of acceptance: Sophomore < Junior
Acceptance of facial expression recording: Sophomore < Junior, Sophomore < Senior

(M = 4.47, SD = 0.61), while juniors reported the highest level of general acceptance (M = 4.70, SD = 0.51). Post hoc multiple comparisons showed that sophomores' acceptance scores were significantly lower than those of juniors, suggesting that students in the earlier stages of their undergraduate education may be less familiar with or confident about the educational integration of mobile technologies.

A similar pattern was observed for acceptance of facial expression recording, where ANOVA also revealed a significant main effect of academic year, F = 9.829, $p < .001$. Once again, sophomores reported the lowest mean score (M = 2.64, SD = 0.50), while juniors (M = 3.00, SD = 0.64) and seniors (M = 3.01, SD = 0.61) reported comparably higher levels of acceptance. Multiple comparison tests identified two statistically significant differences: sophomores' acceptance was significantly lower than that of both juniors and seniors. This finding is particularly important as it highlights a clear developmental trajectory in students' attitudes toward the use of affective technologies. As students progress through university, they may become more accustomed to monitoring and assessment systems in academic contexts, more familiar with the pedagogical rationale behind emotion-sensing tools, or more confident in managing the implications of being recorded. Furthermore, juniors and seniors may be more likely to encounter data privacy policies, ethical discussions, and educational technologies in advanced coursework, fostering more informed judgments about tools such as facial expression analysis. The absence of a significant difference between juniors and seniors suggests that acceptance levels may stabilize after the mid-point of students' academic experience. No significant differences were found across academic years in ethical and fairness concerns, F = 0.706, p = .495. Sophomores (M = 4.27, SD = 0.72), juniors (M = 4.36, SD = 0.64), and seniors (M = 4.20, SD = 0.75) all expressed similarly high levels of ethical concern, indicating that ethical awareness transcends academic seniority. Similarly, conditional acceptance did not significantly differ across groups, F = 1.104, p = .334. This dimension was measured based on students' responses to Q17 after the presentation of ethical safeguard conditions (opt-out rights, use of personal devices, beautification options, and clear explanations, as outlined in Q16). Conditional acceptance scores were high across all year groups: sophomores (M = 4.45, SD = 0.50), juniors (M = 4.55, SD = 0.50), and seniors (M = 4.54, SD = 0.50). This finding suggests that when students are assured of ethical transparency and procedural fairness, they express a strong willingness to participate in facial expression recording, regardless of academic standing.

4.2 Results of Study 2

4.2.1 Consistency of Dominant Emotions Between Initial Impression and Delayed Impression

The cross-tabulation of consistency ensured the systematic and reliable identification of dominant emotional expressions for each participant. A CNN model was used to classify facial expressions into 7 categories, and the dominant emotion for each participant was determined based on the longest duration, calculated using Min-Max normalization (He et al., 2016; Rafique et al., 2022). Table 4.11 displays a cross-tabulation of the dominant facial emotions identified under the Initial Impression (II) and Delayed Impression (DI) conditions. The analysis revealed complete consistency between the two conditions: all 15 participants who were classified as exhibiting a dominant "neutral" expression in the II condition were likewise identified as "neutral" in the DI condition, and all 9 participants whose dominant emotion was classified as "sad" in the II condition were also categorized as "sad" in the DI condition. No discrepancies in emotion categories were observed between the two conditions. These results suggest a high degree of consistency in the expression of dominant facial emotions across II and DI conditions.

The self-reported textual data presented in Table 4.12 further support hypothesis 1, indicating that the majority of participants demonstrated consistent self-reported emotions. Specifically, among the 24 participants, 19 were consistently classified as expressing a dominant emotion of "good" under both the II and DI conditions, while 3 participants were consistently categorized as expressing "dislike."

Table 4.11: Cross-tabulation of facial impressions between initial and delayed impressions.

II / DI	Neutral	Sad	Total
Neutral	15	0	15
Sad	0	9	9
Total	15	9	24

II = Initial Impression; DI = Delayed Impression.

Table 4.12: Cross-tabulation of self-reported texts between initial and delayed impressions.

II / DI	Dislike	Good	Total
Dislike	3	1	4
Good	1	19	20
Total	4	20	24

II = Initial Impression; DI = Delayed Impression.

However, two participants exhibited discrepancies: one participant's dominant emotion shifted from "good" in the II condition to "dislike" in the DI condition, while the other showed the reverse pattern. Compared to facial expression data, the consistency of dominant textual emotions across the two conditions was relatively high, though not perfect. Overall, the majority of participants (22 out of 24) demonstrated consistent dominant emotional expressions between the II and DI conditions. The Cohen's Kappa coefficients and associated significance levels for the consistency of dominant emotions identified from facial expressions and self-reported textual responses are presented in Table 4.13. For the facial expression data, the Kappa coefficient was 1.000 ($p < .001$), indicating perfect agreement between the II and DI conditions (Sim and Wright, 2005). This result corroborates the descriptive findings in Table 4.11 and demonstrates the complete stability of participants' dominant emotional expressions across II and DI conditions. The statistical significance of this coefficient further underscores the robustness of the facial emotion classification procedure employed in the study. In contrast, the Kappa coefficient for the self-reported textual responses was 0.700 ($p = .001$), indicating substantial agreement. Although this consistency remains statistically significant at the $p < .05$ level, the lower Kappa value and higher p-value, relative to the facial expression data, reflect greater variability in participants' self-reported emotional states. As shown in Table 4.12, two participants exhibited shifts in the valence of their dominant emotions across conditions, which may be attributable to the reflective and

Table 4.13: Consistency of dominant emotions between initial and delayed impressions.

Items	kappa	p	sig.
Facial expression	1.000	$< .001$	**
Self-report text	0.700	.001	**

**$p < .01$.

reconstructive nature of written self-reports under the DI condition.

These results highlight an important distinction between automatic, observable emotional responses (facial expression) and subjectively constructed emotional reflections (self-report text). The former, as captured through facial expressions, appear to be more temporally and structurally stable, whereas the latter may be more susceptible to post-experiential cognitive processes, memory recall, and reinterpretation. Nonetheless, a certain degree of consistency between these two forms of emotional expression is maintained within a bounded range.

4.2.2 Comparative Analysis of Initial and Delayed Impressions

4.2.2.1 Comparison Across Groups

The mean difference scores presented in Table 4.14 illustrate changes across five emotional indicators from the Initial Impression (II) to the Delayed Impression (DI) condition for the four experimental groups. The relatively consistent direction and magnitude of change across groups highlight a pronounced temporal effect on emotional expression following a three-day delay.

Specifically, the DI–II mean differences for Facial Emotion Categories (FEC), ranging from 1.33 to 2.00, and Textual Emotion Categories (TEC), ranging from 2.00 to 2.33, were positive for all groups, indicating an increase in the diversity of both facial and textual emotional expressions after the delay. Although the changes in Non-dominant Facial Emotions (NFE) were modest, they followed the same trend, suggesting that participants were more likely to articulate secondary or less dominant emotions upon reflection, with Group 1 showing the greatest shift (0.23). The indices related to textual emotional richness: Total Non-dominant Emotion Words (TNEW), Non-dominant Emotional Density (NED), and Non-dominant Emotional Intensity per Sentence (NEIPS), all demonstrated overall increases from II to DI. Group 2 exhibited the most substantial differences in TNEW (8.00),

Table 4.14: Mean differences of emotional indicators across groups.

Items	Mean(DI - II) Group[1]	Group[2]	Group[3]	Group[4]
FEC	1.83	1.33	2.00	2.00
TEC	2.33	2.17	2.00	2.33
NFE	0.23	0.21	0.17	0.20
TNEW	6.50	8.00	4.83	5.83
NED	0.023	0.027	0.026	0.022
NEIPS	0.59	0.71	0.57	0.42

II = Initial Impression; DI = Delayed Impression; FEC = facial emotion categories; TEC = textual emotion categories; NFE = non-dominant facial emotions; TNEW = total non-dominant emotion words; NED = non-dominant emotion density; NEIPS = non-dominant emotion intensity per sentence.

NED (0.19), and NEIPS (0.71), indicating the strongest delayed emotional response among all groups. While these descriptive results have not yet undergone statistical testing, the consistency and magnitude of change suggest meaningful differences that warrant further analysis through one-way ANOVA. These findings support the assumption of sufficient between-group variance, thereby justifying the use of inferential statistics in subsequent analyses to determine whether group-level differences in emotional amplification are statistically significant.

Table 4.15 presents the results of a one-way analysis of variance (ANOVA) conducted to examine whether the observed mean differences between II and DI conditions across five emotional indicators varied significantly among the four experimental groups. At the conventional alpha level of .05, none of the comparisons reached statistical significance, with p-values ranging from .227 to .959.

The analyses of FEC: $F(3,10) = 1.02$, $p = .422$, TEC: $F(3,11) = 0.10$, $p = .959$, and NFE: $F(3,10) = 0.41$, $p = .753$ indicated minimal between-group variance. Similarly, the text-based indicators, TNEW: $F(3,11) = 1.54$, $p = .259$, and NEIPS: $F(3,11) = 1.69$, $p = .227$, also failed to demonstrate significant group-level differences. Effect size estimates (r), calculated as the square root of η^2, further supported the absence of strong group effects, with values ranging from $r = .138$ to $r = .465$. These values suggest small to moderate effect sizes (Fritz et al., 2012).

The non-significant ANOVA results have important methodological implications for subsequent statistical analyses, particularly for the use of paired sample t-tests to assess within-subject changes between the II and DI conditions at the aggregated level. A key concern in performing t-tests

Table 4.15: One-way ANOVA results on differences in initial and delayed impressions across groups.

Items	f	p	r
FEC	(3, 10) = 1.02	.422	0.286
TEC	(3, 11) = 0.10	.959	0.138
NFE	(3, 10) = 0.41	.753	0.259
TNEW	(3, 11) = 1.54	.259	0.465
NED	(3, 10) = 0.71	.568	< 0.1
NEIPS	(3, 11) = 1.69	.227	0.409

II = Initial Impression; DI = Delayed Impression; FEC = facial emotion categories; TEC = textual emotion categories; NFE = non-dominant facial emotions; TNEW = total non-dominant emotion words; NED = non-dominant emotion density; NEIPS = non-dominant emotion intensity per sentence.
r = $\sqrt{\eta^2}$.

on aggregated data from multiple experimental groups is whether any specific group exhibits disproportionately high or low difference scores, thereby introducing random variance or bias into the overall results. However, the absence of statistically significant group effects, along with the observation of small to moderate effect sizes, suggests that emotional responses across groups were sufficiently homogeneous (Liu and Wang, 2021). This homogeneity provides statistical justification for aggregating groups in the subsequent paired-sample t-tests , as it indicates that no single group unduly influenced the direction or magnitude of the observed differences between II and DI conditions. Moreover, the ANOVA results enhance the internal validity of the paired-sample t-test by mitigating concerns regarding random group-level artifacts. Had the ANOVA revealed significant group effects, it would have suggested the presence of systematic bias associated with group assignment, such as differences in video content or classroom dynamics, that could compromise the interpretability of aggregated results. Therefore, the lack of significant group differences supports the conclusion that any observed changes between the II and DI conditions can be attributed to the temporal dimension of emotional processing, rather than confounding factors specific to particular experimental groups.

Particular attention was given to the results of the NED variable. Unlike the other indicators, NED reported an extremely small effect size, with an r value of less than 0.1. According to standard conventions, an r value below 0.1 is considered negligible (Fritz et al., 2012), indicating a lack of meaningful variance across groups. In light of this finding, and to maintain statistical rigor and ensure valid inference, the study conducted a separate

Table 4.16: Summary statistics prior to paired sample t-test between initial and delayed impressions.

	II		DI	
Items	Mean	S.D.	Mean	S.D.
FEC	4.42	0.93	6.21	0.72
TEC	2.63	0.97	4.83	0.91
NFE	0.20	0.12	0.41	0.12
TNEW	3.63	2.86	9.92	4.80
NEIPS	0.44	0.31	1.01	0.41

II = Initial Impression; DI = Delayed Impression; FEC = facial emotion categories; TEC = textual emotion categories; NFE = non-dominant facial emotions; TNEW = total non-dominant emotion words; NEIPS = non-dominant emotion intensity per sentence.

non-parametric analysis using the Wilcoxon signed rank test for the NED variable. By applying the Wilcoxon test specifically to NED, this analysis preserved the methodological integrity of Study 2 and ensured that the conclusions drawn regarding non-dominant emotion density under the II and DI conditions were based on appropriate statistical reasoning.

4.2.2.2 Comparison of Emotional Category Diversity and Intensity of Non-Dominant Emotions Between Initial and Delayed Impressions

Table 4.16 provides critical empirical support for the subsequent paired sample t-tests and establish a quantitative foundation for testing the central hypothesis of Study 2. The FEC and TEC were derived using a CNN-based emotion recognition model, which classified each participant's facial or textual expression into one of seven discrete emotion categories: anger, disgust, fear, happy, sad, surprise, and neutral. The number of distinct emotional categories expressed by each participant under both conditions was calculated. For FEC, the average number of facial emotion categories increased from 4.42 (SD = 0.93) in the II condition to 6.21 (SD = 0.72) in the DI condition. For TEC, the mean number of textual emotion categories increased from 2.63 (SD = 0.97) to 4.83 (SD = 0.91). These substantial increases suggest an elevated degree of emotional diversity in the delayed condition, supporting the first part of Hypothesis 2 that deeper cognitive-affective processing over time leads to greater emotional variety in DI responses.

The remaining indicators provide evidence related to the intensity and complexity of non-dominant emotional responses, an essential component of Hypothesis 2. NFE refers to the sum of all emotional states expressed by participants excluding the dominant (longest-duration) emotion. The average NFE increased from 0.20 (SD = 0.12) in the II condition to 0.41 (SD = 0.12) in the DI condition, indicating a significant rise in emotional complexity or ambivalence over time. This suggests that participants were better able to reflect on their initial reactions and express more nuanced emotional states that may not have been immediately accessible. Text-based indicators reinforced this trend. The TNEW, quantifying the number of affective terms used beyond the dominant emotion, nearly tripled from a mean of 3.63 (SD = 2.86) in the II condition to 9.92 (SD = 4.80) in the DI condition. Similarly, NEIPS, an index of expressive force, showed a significant increase from 0.44 (SD = 0.31) in the II condition to 1.01 (SD = 0.41) in the DI condition, further emphasizing the increased and more frequent use of emotional language in delayed textual responses.

Collectively, the descriptive patterns observed in Table 4.16 offer robust preliminary support for the hypotheses proposed in Study 2. First, emotional diversity, both facial and textual, was substantially greater in the DI condition than in II, validating the core assumption that temporal distance in TBMA exposure facilitates broader emotional engagement. Second, non-dominant emotions were significantly amplified under the DI condition. Furthermore, as demonstrated in the prior one-way ANOVA (Table 4.15), the absence of significant between-group differences justifies the statistical decision to aggregate all four experimental groups for within-subject comparison. This homogeneity strengthens the generalizability of the findings and underscores the primacy of temporal structure, rather than video content or participant subgroup, as the central factor driving emotional change in this study. The summary statistics provided in Table 4.16 serve both as a diagnostic and validation step in the inferential analysis of emotional responses to TBMA. They establish a clear empirical basis for conducting paired-sample t-tests and provide compelling preliminary descriptive evidence in support of the study's methodological framework and hypotheses.

Table 4.17 presents the results of the paired sample t-tests, all of which yielded statistically significant t-values at the $p < .001$ level (Blair and Higgins, 1985), offering strong evidence of substantial differences between the II and DI conditions. The magnitude of these temporal effects was further underscored by consistently large effect sizes, with all Cohen's d values exceeding 1.3 (Fritz et al., 2012). These findings provide robust statistical validation for Hypothesis 2, which posits that DI would elicit greater emotional diversity, and DI would be associated with more intense and nuanced emotional expression, particularly in relation to non-dominant emotions.

Table 4.17: Comparison of paired sample t-test between initial impression and delayed impressions.

Items	$t_{23}(II - DI)$	p	d	Sig.
FEC	-8.98	$< .001$	1.834	**
TEC	-10.60	$< .001$	2.163	**
NFE	-10.59	$< .001$	2.173	**
TNEW	-12.19	$< .001$	2.489	**
NEIPS	-11.06	$< .001$	2.260	**

II = Initial Impression; DI = Delayed Impression; FEC = facial emotion categories; TEC = textual emotion categories; NFE = non-dominant facial emotions; TNEW = total non-dominant emotion words; NEIPS = non-dominant emotion intensity per sentence. d=Cohen's d, ** $p < .01$.

Participants expressed a significantly greater number of emotion categories under the DI condition than under the II condition, both in facial expressions (FEC: t(23) = -8.98, $p < .001$, d = 1.834) and textual reflections (TEC: t(23) = -10.60, $p < .001$, d = 2.163). These large effect sizes suggest that the three-day delay, combined with cognitive recall mechanisms and visual cues from still images, meaningfully enriched participants' emotional responses to TBMA stimuli.

The second half of the Table 4.17 focuses on non-dominant emotional indicators, revealing deeper and more complex emotional expression in the DI condition. NFE were significantly higher in the DI condition (t(23) = -10.59, $p < .001$, d = 2.173), indicating that participants were more likely to express secondary or previously suppressed emotional states after delayed exposure. The TNEW measure showed the largest effect size among all indicators (t(23) = -12.19, $p < .001$, d = 2.489), highlighting a substantial increase in the use of affective vocabulary linked to less dominant emotional states during delayed reflection. Similarly, NEIPS rose sharply (t(23) = -11.06, $p < .001$, d = 2.260), further reinforcing the interpretation that DI reflections were not only longer and more affectively rich, but also more emotionally saturated at the sentence level.

These findings collectively indicate a deepening of emotional expression after a three-day interval. The descriptive statistics reported in Table 4.16 also mirrored this pattern, with substantial increases in all five indicators from II to DI. For example, mean FEC increased from 4.42 to 6.21, TEC from 2.63 to 4.83, and similar upward shifts were observed for NFE (0.20 to 0.41), TNEW (3.63 to 9.92), and NEIPS (0.44 to 1.01). These descriptive changes suggest that the observed effects are not merely statistical artifacts but reflect consistent and substantive emotional developments over time. The

Table 4.18: Nonparametric test results of NED.

II - DI	N	Mean Rank	Sum of Ranks
Negative Ranks	24[a]	12.50	300.00
Positive Ranks	0[b]		
Ties	0[c]		
Wilcoxon Signed Ranks Test		Z	-4.286
		p	$< .001$

NED = non-dominant emotion density; II = initial impression; DI = delayed impression
[a] II < DI; [b] II > DI; [c] II = DI.
Z: based on positive ranks.

robustness of these results can be attributed to the rigorous preprocessing procedures detailed in the methodology section. For facial emotion analysis, the CNN model classified and quantified emotion from each video frame using 1-fps sampling, optimizing computational efficiency while preserving core emotional signals. Despite variations in video duration, normalization and standardization procedures enabled fair cross-condition comparisons. In the case of textual analysis, participants' self-reported reflections were analyzed using a Chinese emotion Dictionaries, from which non-dominant emotional terms were extracted, and higher-order indicators such as TNEW and EIPS were calculated (Liu, 2022; Mohammad, 2016, 2021). These preprocessing strategies ensured that all emotion metrics were standardized, interpretable, and sensitive to subtle affective changes, thereby enhancing the reliability and replicability of the statistical findings.

Table 4.18 presents the results of the Wilcoxon signed rank test conducted for the variable NED. This non-parametric test was selected as an alternative to the paired-sample t-test due to the extremely low variance observed across experimental groups (r < 0.1, as shown in Table 4.15).

As displayed in the Table 4.18, all 24 participants reported higher NED values under the DI condition compared to the II condition, resulting in 24 negative ranks, 0 positive ranks, and 0 ties. The complete dominance of negative ranks indicates a consistent pattern across all participants: the emotional density in their delayed self-reports was uniformly greater than in their initial responses. The Wilcoxon test yielded a Z-value of -4.286 and a p-value less than .001, indicating that the increase in non-dominant emotion density from II to DI was not only systematic but also highly statistically significant. This finding offers strong non-parametric confirmation of the trends observed in the paired-sample t-test analyses. NED measures the proportion of non-dominant emotional vocabulary within the total lexical content of participants' reflections, capturing the subtlety and richness of their affective responses. The significant increase in NED under the DI condition reinforces the earlier hypothesis that delayed exposure to TBMA not only expands the breadth of emotional engagement, but also makes emotions more profound.

In conclusion, the results in Table 4.17 and Table 4.18 provide compelling and statistically robust evidence that temporally structured TBMA exposure significantly enhances both the diversity and intensity of students' emotional responses. The significant increases across all six indicators, confirm that delayed engagement fosters a deeper, more reflective, and more nuanced emotional interpretation of media art. These findings not only support the hypotheses of Study 2, but also provide critical guidance for subsequent case studies, suggesting that the integration of emotion recognition technologies into well-designed curricular timelines can meaningfully enrich students' emotional and aesthetic engagement in contemporary media arts education.

Chapter 5

Results of the Relationship Between Emotional Reappraisal and Art Appreciation (Study 3)

5.1 Consistency in Dominant Emotions Between Initial and Delayed Impressions

To explore the temporal consistency of students' emotional responses, this study employed a dual-modal approach combining facial expression recognition with self-reported textual emotion analysis. Table 5.1 presents the cross-tabulated distribution of dominant emotional categories across facial and textual modalities, comparing responses from the Initial Impression (II) and Delayed Impression (DI) sessions.

In the upper section of the table, facial emotions are categorized into three groups: happy, neutral and sad. Among the 29 participants, the majority (n = 20) showed neutral facial expressions under conditions II and DI. Three participants consistently expressed happy, while four maintained sadness across both II and DI, resulting in a total of 27 cases of facial emotion consistency. Only two participants showed changes: one shifted from neutral to happy, and another from sad to neutral. The lower section of Table 5.1 reflects the analysis of textual emotions based on the affective vocabulary used in participants' written reflections, categorized as "dislike" and "good." Here, the consistency rate was slightly lower but remained high. Of the 14 participants who initially expressed "dislike," 12 remained in that category in the DI session. Similarly, 12 out of 15 participants who initially expressed "good" maintained the same classification after the delay. Only five participants switched categories: three moved from "dislike" to "good," while two shifted in the opposite direction.

In this case study, Table 5.2 reveals a notable shift in week 2 compared to the previous week 1. While week 1 was dominated by the prevalence of neutral facial expressions, week 2 reflected a stronger engagement with more

Table 5.1: Cross-tabulation of facial impression and self-report text between initial and delayed impressions in week 1.

II^F / DI^F	Happy	Neutral	Sad	Total
Happy	3	1	0	4
Neutral	0	20	0	20
Sad	0	1	4	5
Total	3	22	4	29
II^T / DI^T	Dislike	Good	Total	
Dislike	12	3	15	
Good	2	12	14	
Total	14	15	29	

F = facial impression; T = self-report text.
II = Initial Impression; DI = Delayed Impression.

distinct emotions, particularly anger, disgust, and fear. Eight participants consistently displayed angry expressions, twelve maintained disgust, and six showed fear. Only a few deviations were observed: one participant shifted from fear to neutral, and two exhibited sad in the DI condition that was not present during the II condition. This suggests that the Time-Based Media Art (TBMA) content presented in week 2 have been more emotionally provocative than that of week 1. In the textual dimension, the overall pattern remained relatively consistent with that of week 1. Most participants who reported a "dislike" emotion in their self-reflections (n = 13) retained this classification during delayed reflection, while those who initially reported a "good" impression (n = 10) showed a similar trend. A small number of participants exhibited changes in their textual emotional evaluations: two shifted from "good" to "dislike," and four from "dislike" to "good."

The results from week 3, as illustrated in Table 5.3, indicate that the three dominant emotional states in both the II and DI conditions were happy, neutral, and sad. Among the nine participants categorized as "happy" in the II condition, seven maintained this classification in the DI condition, while two transitioned to "neutral." Notably, none of the participants who initially displayed happiness shifted to sadness, suggesting emotional stability in positive responses. For neutral, all 16 participants retained the same label across both sessions, reflecting the highest level of consistency among all emotional classifications. In contrast, participants initially categorized as "sad" (n = 4) also demonstrated perfect consistency, remaining in the same emotional category over time. Regarding text-based emotional self-reports, among the 13 participants who reported a "dislike" response in the II condition, nine maintained the same classification in the DI condition,

Table 5.2: Cross-tabulation of facial impression and self-report text between initial and delayed impressions in week 2.

II^F / DI^F	Angry	Disgust	Fear	Neutral	Sad	Total
Angry	8	0	0	0	0	8
Disgust	0	12	0	0	0	12
Fear	0	0	6	1	1	8
Neutral	0	0	0	0	1	1
Total	8	12	6	1	2	29
II^T / DI^T	Dislike	Good	Total			
Dislike	13	4	17			
Good	2	10	12			
Total	15	14	29			

F = facial impression; T = self-report text.
II = Initial Impression; DI = Delayed Impression.

while four shifted to a "good" evaluation. Conversely, of the 16 participants who initially reported "good" impressions, fourteen remained stable, with only two transitioning to "dislike."

As part of the four-week study on engagement with TBMA, Table 5.4 showed the final trends in cumulative exposure and its relationship to participants' changing emotional resonance. Among the 29 participants, 28 maintained the same dominant facial emotion classification across both the II and DI conditions. Specifically, all three participants classified as angry in the II remained angry in the DI. The two participants categorized as fearful also retained their classification. Of the eight participants labeled as happy in II, all continued to be classified as happy in DI. Notably, within neutral, 11 of the 12 participants maintained a neutral classification, with only one transitioning to sad. The single participant marked as sad in the II remained so in DI. Compared to facial expressions, self-reported textual emotions showed slightly greater variability. Of the 17 participants who reported a "dislike" emotion in II, 14 retained this classification in DI, while 3 shifted to "good." Additionally, among the 12 participants who initially reported "good" impressions in II, 11 remained consistent, with only one shifting to "dislike." Although most participants displayed consistent emotional responses across the two conditions, these moderate shifts suggest that reflective judgment remained an active cognitive process, even in the final stage of the case study.

The cross-tabulation results from week 1 through week 4 demonstrate a high degree of emotional convergence and stability. Facial expression data showed greater consistency, which may reflect participants' habituation to the

Table 5.3: Cross-tabulation of facial impression and self-report text between initial and delayed impressions in week 3.

II^F / DI^F	Happy	Neutral	Sad	Total
Happy	7	2	0	9
Neutral	0	16	0	16
Sad	0	0	4	4
Total	7	18	4	29
II^T / DI^T	Dislike	Good	Total	
Dislike	9	4	13	
Good	2	14	16	
Total	11	18	29	

F = facial impression; T = self-report text.
II = Initial Impression; DI = Delayed Impression.

experimental conditions or repeated affective responses to recurring TBMA aesthetics. Meanwhile, textual self-reports exhibited ongoing reflective engagement, highlighting the complementary nature of multimodal emotion assessment. These findings support the methodological robustness of integrating facial and textual data and reinforce the utility of delayed impressions for capturing longitudinal emotional dynamics in the appreciation of TBMA.

Following the descriptive results of the cross-tabulations, Table 5.5 presents Cohen's Kappa coefficients used to evaluate the consistency of dominant emotions between II and DI across four consecutive weeks. These analyzes were conducted separately for facial expression recognition (processed using CNN model) and self-reported textual data (analyzed using Chinese emotion dictionary).

For facial expressions, Kappa coefficients were consistently high across all four weeks: 0.843 (week 1), 0.851 (week 2), 0.878 (week 3), and peaking at 0.953 in week 4. According to the widely accepted interpretation guidelines proposed by Landis and Koch (Landis and Koch, 1977), values between 0.81 and 1.00 indicate almost perfect agreement. These results strongly suggest that facial recognition captured highly stable emotional patterns over time. The increasing trend in Kappa values from week 1 to week 4 indicates a progressive consolidation of emotional responses, possibly reflecting deeper cognitive-affective integration as participants were repeatedly exposed to TBMA stimuli.

In contrast, Kappa values for self-reported textual responses demonstrated moderate to substantial agreement, but consistently fell below those of the facial data. Specifically, values ranged from 0.576 (week 3) to 0.722

Table 5.4: Cross-tabulation of facial impression and self-report text between initial and delayed impressions in week 4.

II^F / DI^F	Angry	Fear	Happy	Neutral	Sad	Total
Angry	3	0	0	0	0	3
Fear	0	2	0	0	0	2
Happy	0	0	8	0	0	8
Neutral	0	0	0	11	1	12
Sad	0	0	0	0	4	1
Total	3	2	8	11	5	29
II^T / DI^T	Dislike	Good	Total			
Dislike	14	3	17			
Good	1	11	12			
Total	15	14	29			

F = facial impression; T = self-report text.
II = Initial Impression; DI = Delayed Impression.

Table 5.5: Consistency of dominant emotions between initial and delayed impressions in four weeks.

$\text{Items}^{\text{week 1}}$	kappa	p	sig.
Facial expression	0.843	$< .001$	**
Self-report text	0.656	$< .001$	**
$\text{Items}^{\text{week 2}}$			
Facial expression	0.851	$< .001$	**
Self-report text	0.584	.002	**
$\text{Items}^{\text{week 3}}$			
Facial expression	0.878	$< .001$	**
Self-report text	0.576	.002	**
$\text{Items}^{\text{week 4}}$			
Facial expression	0.953	$< .001$	**
Self-report text	0.722	$< .001$	**

$^{**}p < .01$.

(week 4), with week 1 and week 2 showing values of 0.656 and 0.584, respectively. These figures correspond to moderate (0.41–0.60) to substantial (0.61–0.80) agreement. The discrepancy between facial and textual consistency highlights the more interpretive and reflective nature of text-based responses. Participants may have reinterpreted or elaborated on their emotional experiences over time, particularly in the DI condition, where recall and reflective self-reporting were involved. Therefore, the relatively lower Kappa values for text should not be interpreted as measurement error across modalities, but rather as an indication of the dual nature of emotion, both as immediate reaction and as reconstructed, cognitively filtered narrative.

5.2 Differences in Non-dominant Emotions Between Initial and Delayed Impressions

Following the assessment of dominant emotion consistency through Cohen's Kappa coefficients in both facial and textual modalities, this case study further examined whether participants' emotions became more diverse and intense under the DI condition across each of the four weeks. Table 5.6 provides a comparative summary of descriptive statistics (means and standard deviations) for five emotional indicators under the Initial Impression (II) and DI conditions: Facial Emotion Categories (FEC), Textual Emotion Categories (TEC), Non-Focal Emotions (NFE), Total Non-Dominant Emotion Words (TNEW), and Non-Dominant Emotion Intensity Per Sentence (NEIPS).

Table 5.6: Comparative descriptive statistics of emotional indicators between initial and delayed impressions across four weeks.

$\text{Items}^{\text{week 1}}$	Mean^{II}	S.D.^{II}	Mean^{DI}	S.D.^{DI}
FEC	4.52	0.74	4.93	0.72
TEC	3.45	0.87	4.52	0.99
NFE	0.31	0.11	0.33	0.12
TNEW	4.52	2.63	10.31	4.29
NEIPS	0.62	0.36	1.05	0.51
$\text{Items}^{\text{week 2}}$				
FEC	4.34	0.55	5.17	0.47
TEC	3.14	0.83	3.90	0.82
NFE	0.41	0.12	0.52	0.07
TNEW	4.93	3.65	8.76	3.94

Table 5.6 (continued)

NEIPS	0.39	0.29	0.61	0.33
Items$^{\text{week 3}}$				
FEC	4.07	0.80	5.17	0.47
TEC	3.45	0.69	4.45	0.99
NFE	0.14	0.07	0.23	0.15
TNEW	4.90	2.53	9.10	3.72
NEIPS	0.41	0.26	0.62	0.30
Items$^{\text{week 4}}$				
FEC	4.28	0.65	5.83	0.71
TEC	3.34	0.72	4.59	0.63
NFE	0.28	0.07	0.42	0.11
TNEW	4.28	2.48	16.66	4.55
NEIPS	0.61	0.35	1.22	0.34

Note. II = Initial Impression; DI = Delayed Impression; FEC = facial emotion categories; TEC = textual emotion categories; NFE = non-dominant facial emotions; TNEW = total non-dominant emotion words; NEIPS = non-dominant emotion intensity per sentence.

A clear and consistent trend is evident in Table 5.6: for all indicators and across all weeks, DI values are higher than their corresponding II values. As shown in Table 5.6, FEC, which reflects facial emotional diversity detected through CNN analysis, steadily increased from II to DI. week 1 shows a moderate rise from M = 4.52 (II) to M = 4.93 (DI), while week 4 demonstrates a more substantial increase from M = 4.28 to M = 5.83. This progressive expansion of facial emotion categories indicates that participants' delayed emotional responses became increasingly diverse over time. A similar trend was observed for TEC, which captures the variety of emotional categories expressed in participants' self-reported texts. In all weeks, the DI mean exceeded the II mean: week 1 (3.45 to 4.52), week 2 (3.14 to 3.90), week 3 (3.44 to 4.45), and week 4 (3.34 to 4.59), with the most pronounced increase occurring in week 1. Turning to non-dominant emotion indicators, NFE, TNEW, and NEIPS all showed increases from II to DI, supporting the hypothesis that delayed reflection enhances emotional intensity. For instance, TNEW surged from 4.28 to 16.66 in week 4. Similar trends were noted in earlier weeks, such as week 1 (6.63 to 10.31) and week 2 (4.93 to 8.76), suggesting that reflection over time expands the articulation and breadth of emotional vocabulary. NEIPS also increased across all weeks, especially in week 4, where the average rose from 0.61 to 1.22, reinforcing the interpretation that DI not only contain more emotional language but also convey greater emotional intensity per sentence. It is important to emphasize that, due to preprocessing constraints in facial emotion data, NFE values were relatively lower than those of other indicators. Nonetheless, NFE consistently increased from II to DI across all weeks. For example, in week 3, NFE rose from 0.28 to 0.42, aligning with the broader trend that DI conditions elicited more nuanced and emotionally layered expressions.

Following the descriptive statistics on mean differences observed from week 1 through week 4, Table 5.7 presents a detailed summary of the paired sample t-test results comparing emotional indicators between II and DI across four consecutive weeks.

Beginning with week 1, statistically significant increases were observed in FEC (t(28) = -2.857, p = .008, d = 0.531) and TEC (t(29) = -5.574, p < .001, d = 1.035), indicating that participants expressed a broader range of both facial and textual emotions under the DI condition. While the effect size for FEC falls within the moderate range (Fritz et al., 2012), the large effect size for TEC suggests a meaningful shift in emotional clarity over time through cognitive elaboration. Likewise, a statistically significant but smaller effect was observed for NFE (t(28) = -2.188, p = .037, d = 0.400). Substantial effect sizes were reported for TNEW (t(28) = -8.550, p < .001, d = 1.588) and NEIPS (t(28) = -6.278, p < .001, d = 1.130), supporting the hypothesis that delayed reflection increases emotional density and expressive intensity in self-reported texts.

Table 5.7: Paired sample T-test results comparing emotional indicators between initial and delayed impressions across four weeks.

Items$^{\text{week 1}}$	$t_{28}(II - DI)$	p	d	Sig.
FEC	-2.857	.008	0.531	**
TEC	-5.574	< .001	1.035	**
NFE	-2.188	.037	0.400	*
TNEW	-8.550	< .001	1.588	**
NEIPS	-6.278	< .001	1.130	**
Items$^{\text{week 2}}$				
FEC	-6.272	< .001	1.165	**
TEC	-5.925	< .001	1.102	**
NFE	-4.270	< .001	0.802	**
TNEW	-7.419	< .001	1.378	**
NEIPS	-4.672	< .001	0.863	**
Items$^{\text{week 3}}$				
FEC	-7.273	< .001	1.350	**
TEC	-4.887	.001	0.907	**

Table 5.7 (continued)

NFE	-3.608	< .001	0.669	**
TNEW	-7.006	< .001	1.301	**
NEIPS	-3.376	.002	0.626	**
Items$^{\text{week 4}}$				
FEC	-9.610	< .001	1.784	**
TEC	-6.534	< .001	1.213	**
NFE	-5.310	< .001	0.986	**
TNEW	-14.637	< .001	2.718	**
NEIPS	-6.856	< .001	1.274	**

Note. II = Initial Impression; DI = Delayed Impression; FEC = facial emotion categories; TEC = textual emotion categories; NFE = non-dominant facial emotions; TNEW = total non-dominant emotion words; NEIPS = non-dominant emotion intensity per sentence. d=Cohen's *d*; * $p < .05$.; ** $p < .01$..

By week 2, the patterns observed in week 1 became more robust. All five indicators showed statistically significant differences between II and DI at the $p < .001$ level, with effect sizes exceeding 0.80. FEC (d = 1.165) and TEC (d = 1.102) continued to reflect heightened emotional diversity, while NFE (d = 0.802), TNEW (d = 1.378), and NEIPS (d = 0.863) emphasized the sustained deepening of emotional expression over time. week 3 demonstrated similar strong patterns, with FEC (t(28) = -7.273, $p < .001$, d = 1.350) and TEC (t(29) = -4.887, $p < .01$, d = 0.907) again confirming ongoing emotional diversification. The result for NFE (t(28) = -0.20, $p = .001$, d = 0.669) also supported a statistically significant difference, with a moderate-to-large effect size. TNEW and NEIPS both remained significant, although NEIPS showed a slightly smaller effect size this week, possibly reflecting contextual or TBMA content-related variation. By week 4, the differences between II and DI peaked. FEC (t(28) = -9.610, $p < .001$, d = 1.784) and TEC (t(29) = -6.534, $p < .01$, d = 1.213) exhibited very large effect sizes, indicating that participants' capacity for emotional differentiation had significantly advanced. Similarly, NFE (d = 0.986), TNEW (d = 2.718), and NEIPS (d = 1.274) confirmed the strongest changes recorded across the entire study. The especially large effect size for TNEW suggests that students not only used a greater volume of emotional vocabulary but also expressed a more complex emotional range following delayed reflection.

To further investigate participants' emotional responses under the II and DI conditions across four consecutive weeks, Table 5.8 presents the results of the Wilcoxon signed rank test applied to Non-Dominant Emotion Density (NED). This analysis supplements the paired sample t-test results reported in Table 5.7. Given the insufficient effect sizes observed in preliminary variance analyses of NED, the data did not meet the assumptions required for parametric tests such as the paired t-test. Consequently, the Wilcoxon signed-rank test was employed to evaluate whether participants exhibited a statistically significant increase in NED following a three-day delay.

Table 5.8: Nonparametric test results of NED in four weeks.

II - DI$^{\text{week 1}}$	N	Mean Rank	Sum of Ranks
Negative Ranks	26^{a}	15.73	409.00
Positive Ranks	3^{b}	8.67	26.00
Ties	0^{c}		
II - DI$^{\text{week 2}}$			
Negative Ranks	24^{a}	15.63	375.00
Positive Ranks	5^{b}	12.00	60.00
Ties	0^{c}		
II - DI$^{\text{week 3}}$			
Negative Ranks	25^{a}	15.12	378.00
Positive Ranks	4^{b}	14.25	57.00
Ties	0^{c}		
II - DI$^{\text{week 4}}$			
Negative Ranks	28^{a}	14.71	412.00
Positive Ranks	1^{b}	23.00	23.00
Ties	0^{c}		
Wilcoxon Signed Ranks Test$^{\text{week 1}}$		Z	-4.141
		p	$< .001$
Wilcoxon Signed Ranks Test$^{\text{week 2}}$		Z	-3.406
		p	.001
Wilcoxon Signed Ranks Test$^{\text{week 3}}$		Z	-3.471
		p	.001
Wilcoxon Signed Ranks Test$^{\text{week 4}}$		Z	-4.206
		p	$< .001$

NED = non-dominant emotion density; II = initial impression; DI = delayed impression
^{a} II < DI; ^{b} II > DI; ^{c} II = DI.
Z: based on positive ranks.

Across all four weeks, the Wilcoxon tests consistently indicated statistically significant differences in NED between the II and DI conditions, with all p-values less than or equal to .001. These results confirm that the delayed reflection condition elicited a higher density of non-dominant emotional content than the initial response condition. Specifically, week 1 recorded 26 negative ranks and only 3 positive ranks (Z = -4.141, $p < .001$); week 2 showed 24 negative ranks and 5 positive ranks (Z = -3.406, $p = .001$); week 3 indicated 25 negative ranks and 4 positive ranks (Z = -3.471, $p = .001$); and week 4 revealed 28 negative ranks and only 1 positive rank (Z = -4.206, $p =< .001$). Notably, there were no ties in any of the four Wilcoxon analyses, that is, no participants received identical NED scores in both the II and DI conditions. This absence of ties strengthens the interpretive weight of the results, as it indicates that all participants demonstrated a measurable and directional shift in emotional expression density. These findings underscore the sensitivity of the NED measure to experimental manipulation and support the appropriateness of the DI condition as a reflective task.

5.3 Emotional Differences Predicting Changes in Art Understanding, Interest, and Liking

To address RQ2, difference-score regression analyses were conducted to explore how changes in emotional indicators predicted changes in art understanding over four consecutive weeks within the classroom setting of Study 3.

In each weekly (Table 5.9), the Constant term represents the predicted mean change in art understanding when all emotional predictors are held at zero, essentially capturing the baseline difference in understanding that cannot be explained by the emotional indicators. In contrast, the B, SE, and β coefficients of the emotional predictors quantify how much variation in Understanding can be attributed to changes in specific emotional indicators. Therefore, the focus of interpretation lies in the direction, magnitude, and significance of these emotional predictors, while the constant mainly provides a reference point for the intercept (Zou et al., 2003). The regression models revealed a dynamic but coherent pattern. Across the four weeks, NFE and NED consistently showed significant and positive relationships with art understanding, whereas the other indicators (FEC, TEC, TNEW, and NEIPS) displayed more irregular or week-specific influences.

Table 5.9: Results of difference-score regression analysis between emotional indicators and art understanding.

Predictors$^{\text{week 1}}$	B	SE	Beta	t	p
(Constant)	0.872	0.258		3.377	.002
FEC	0.433	0.297	0.271	1.460	.156
(Constant)	0.867	0.340		2.555	.017
TEC	0.172	0.230	0.143	0.748	.416
(Constant)	0.806	0.224		3.605	.001
NFE	24.209	8.413	0.484	2.878	.008**
(Constant)	0.495	0.431		1.150	.260
TNEW	0.096	0.063	0.281	1.519	.140
(Constant)	0.500	0.319		1.570	.128
NED	33.484	14.273	0.411	2.346	.027*
(Constant)	0.818	0.335		2.306	.029
NEIPS	0.537	0.614	0.166	0.875	.389
Predictors$^{\text{week 2}}$					
(Constant)	0.500	0.347		1.442	.161
FEC	0.250	0.320	0.149	0.781	.442
(Constant)	0.724	0.339		2.134	.042
TEC	-0.023	0.334	-0.013	-0.068	.946
(Constant)	0.273	0.258		1.057	.300
NFE	3.999	1.494	0.458	2.676	.012*
(Constant)	0.078	0.360		0.216	.831
TNEW	0.164	0.077	0.382	2.149	.041*
(Constant)	0.314	0.255		1.233	.228
NED	43.750	17.197	0.440	2.544	.017*
(Constant)	0.349	0.283		1.223	.228
NEIPS	1.659	0.868	0.345	1.910	.067
Predictors$^{\text{week 3}}$					
(Constant)	0.262	0.238		1.009	.281
FEC	-0.081	0.175	-0.809	-0.465	.646
(Constant)	-0.070	0.179		-0.393	.698
TEC	0.243	0.121	0.359	2.000	.056
(Constant)	0.445	0.143		3.116	.004
NFE	-2.862	0.845	-0.546	-3.387	.002**
(Constant)	-0.348	0.197		-1.767	.089
TNEW	0.124	0.037	0.537	3.311	.003**
(Constant)	-0.225	0.137		-1.638	.113

Table 5.9 (continued)

NED	42.855	9.437	0.658	4.541	$< .001$***
(Constant)	-0.097	0.136		-0.711	.483
NEIPS	1.317	0.359	0.577	3.672	.001**
Predictors$^{\text{week 4}}$					
(Constant)	0.468	0.333		1.406	.171
FEC	0.082	0.188	0.084	0.436	.666
(Constant)	0.655	0.255		2.563	.016
TEC	-0.048	0.160	-0.058	-0.302	.765
(Constant)	0.209	0.203		1.030	.312
NFE	2.819	1.048	0.460	2.691	.012*
(Constant)	-0.337	0.434		-0.777	.444
TNEW	0.075	0.033	0.402	2.285	.030*
(Constant)	0.078	0.255		0.304	.763
NED	19.638	7.948	0.429	2.471	.020*
(Constant)	0.212	0.247		0.860	.398
NEIPS	0.625	0.318	0.353	1.961	.060

Note. FEC = facial emotion categories; TEC = textual emotion categories; NFE = non-dominant facial emotions; TNEW = total non-dominant emotion words; NED = non-dominant emotion density; NEIPS = non-dominant emotion intensity per sentence.
The dependent variable was the difference in mean art understanding scores between the Delayed Impression (DI) and Initial Impression (II) sessions. The independent variables were the difference scores of six emotional indicators, calculated as DI minus II for each: FEC, TEC, NFE, TNEW, NED, and NEIPS.
* $p < .05$.; ** $p < .01$.; *** $p < .001$.

In week 1 (Table 5.9), two predictors, NFE and NED, showed statistically significant linear relationships with changes in art understanding. NFE exhibited a relatively large standardized coefficient ($\beta = 0.484$, $p = .008$), indicating that the increase in non-dominant facial emotions from II to DI conditions was linearly associated with the improvement in students' mean understanding scores. NED also reached significance ($\beta = 0.411$, $p = .027$), suggesting that greater emotional density in reflective texts predicted a larger positive change in understanding. Although FEC ($\beta = 0.271$, $p = .156$) and TEC ($\beta = 0.143$, $p = .416$) were positively signed, their effects were not statistically significant, indicating that facial and textual emotional diversity contributed weakly to understanding changes.

In week 2 (Table 5.9), the regression patterns displayed stronger linear trends, as the predictive effects of NFE and NED became more robust, and one additional indicator, TNEW, also reached significance. NFE remained a positive predictor ($\beta = 0.458$, $p = .012$), confirming a consistent linear association between the increase of subtle emotional cues and students' interpretive gains. NED continued to predict understanding positively ($\beta = 0.440$, $p = .017$), implying that the compactness and density of emotion expressions in textual reflections were linearly linked to improvements in comprehension. TNEW, representing the total number of non-dominant emotional words, achieved significance ($\beta = 0.382$, $p = .041$), reflecting that an expansion in emotional vocabulary corresponded to a proportional increase in understanding levels. These linear patterns indicated that as participants' emotional articulation became richer and more concentrated, the predicted growth in art understanding followed a corresponding positive slope.

In week 3 (Table 5.9), the linear association between emotional changes and understanding reached its highest strength. NED yielded the largest standardized coefficient across all models ($\beta = 0.658$, $p < .001$). NFE remained significant ($p = .004$), confirming the continuity of its linear influence across sessions. Furthermore, TNEW ($\beta = 0.537$, $p = .003$) and NEIPS ($\beta = 0.577$, $p = .001$) became significant, showing that increases in emotional vocabulary and sentence-level intensity had linear positive relationships with students' interpretive performance.

In week 4 (Table 5.9), the regression coefficients of NFE ($\beta = 0.460$, $p = .012$) and NED ($\beta = 0.429$, $p = .020$) remained significant, reinforcing the temporal stability of their linear predictive power. TEC ($\beta = -0.058$, $p = .765$) and FEC ($\beta = 0.084$, $p = .666$) continued to be nonsignificant, suggesting that emotional category diversity, whether facial or textual, did not exhibit a systematic linear association with understanding improvement. In contrast, TNEW retained a moderate and positive effect ($\beta = 0.402$, $p = .030$).

Across all four weeks, the regression analyses revealed clear linear relationships between changes in non-dominant emotional indicators and gains in art understanding (Table 5.9). The most notable trend occurred for NED in week 3 ($t = 4.541$, $p < .001$), which represented the steepest positive slope observed across all models. This finding underscored that emotional density, an indicator of how compactly emotions were distributed in participants' reflections, served as the strongest linear predictor of cognitive improvement. By contrast, FEC and TEC consistently yielded low t-values (< 2.0) and nonsignificant p-values, reinforcing that superficial emotional variety was not linearly associated with understanding change. The constant terms, ranging from 0.262 to 0.872 and remaining positive across all weeks, indicated an overall upward linear trend in understanding between II and DI, even in the absence of changes in emotional predictors.

As for art interest, Across the four weeks, NED and NEIPS consistently demonstrated significant and positive linear relationships with Interest (Table 5.10). The standardized coefficients (β) of NED rose from 0.432 in week 1 to a peak of 0.794 in week 3 before decreasing slightly to 0.528 in week 4, while NEIPS showed a steady upward trend from 0.373 to 0.690. This pattern reflected a cumulative strengthening of NED and NEIPS as linear predictors of sustained art interest over time.

Table 5.10: Results of difference-score regression analysis between emotional indicators and art interest.

Predictors$^{\text{week 1}}$	B	SE	Beta	t	p
(Constant)	0.346	0.246		1.407	.171
FEC	0.288	0.282	0.193	1.022	.316
(Constant)	0.284	0.317		0.897	.378
TEC	0.170	0.215	0.150	0.790	.436
(Constant)	0.416	0.237		1.751	.091
NFE	4.868	8.932	0.105	0.548	.588
(Constant)	-0.056	0.402		-0.139	.891
TNEW	0.090	0.059	0.282	1.524	.139
(Constant)	-0.075	0.294		-0.256	.800
NED	32.852	13.188	0.432	2.419	.019*
(Constant)	0.026	0.312		-0.083	.934
NEIPS	1.128	0.540	0.373	2.091	.046*
Predictors$^{\text{week 2}}$					
(Constant)	0.110	0.404		0.272	.788
FEC	-0.070	0.373	-0.036	-0.188	.852
(Constant)	0.273	0.387		0.706	.486
TEC	-0.292	0.381	-0.146	-0.768	.449
(Constant)	0.367	0.309		1.186	.246
NFE	3.857	1.790	0.383	2.155	.040*
(Constant)	-1.033	0.368		-2.805	.009
TNEW	0.283	0.078	0.571	3.618	.001**
(Constant)	-0.545	0.267		-2.046	.051
NED	66.562	17.977	0.580	3.703	.001**
(Constant)	-0.562	0.298		-1.884	.070
NEIPS	2.846	0.915	0.513	3.109	.004**
Predictors$^{\text{week 3}}$					
(Constant)	-0.152	0.349		-0.437	.666
FEC	0.271	0.256	0.200	1.060	.299
(Constant)	-0.273	0.259		-1.052	.302
TEC	0.419	0.176	0.417	2.385	.024*

Table 5.10 (continued)

(Constant)	0.280	0.249		1.125	.270
NFE	-1.407	1.476	-0.180	-0.954	.349
(Constant)	-0.933	0.230		-4.057	$<$ .001
TNEW	0.257	0.044	0.749	5.883	$<$ .001***
(Constant)	-0.576	0.165		-3.429	.002
NED	76.860	11.343	0.794	6.776	$<$ .001***
(Constant)	-0.306	0.188		-1.628	.115
NEIPS	2.215	0.495	0.653	4.475	$<$ .001***
Predictors$^{\text{week 4}}$					
(Constant)	-0.331	0.359		-0.920	.365
FEC	0.452	0.203	0.394	2.229	.034*
(Constant)	0.226	0.297		0.761	.453
TEC	0.116	0.186	0.119	0.624	.538
(Constant)	0.150	0.260		0.578	.568
NFE	1.608	1.346	0.224	1.195	.243
(Constant)	-0.917	0.488		-1.880	.071
TNEW	0.104	0.037	0.475	2.807	.009**
(Constant)	-0.373	0.281		-1.330	.195
NED	28.244	8.753	0.528	3.227	.003**
(Constant)	-0.505	0.223		-2.262	.032
NEIPS	1.429	0.288	0.690	4.953	$<$ .001***

Note. FEC = facial emotion categories; TEC = textual emotion categories; NFE = non-dominant facial emotions; TNEW = total non-dominant emotion words; NED = non-dominant emotion density; NEIPS = non-dominant emotion intensity per sentence.
The dependent variable was the difference in mean art interest scores between the Delayed Impression (DI) and Initial Impression (II) sessions. The independent variables were the difference scores of six emotional indicators, calculated as DI minus II for each: FEC, TEC, NFE, TNEW, NED, and NEIPS.
* $p < .05$.; ** $p < .01$.; *** $p < .001$.

In week 1 (Table 5.10), NED (B = 32.85, β = .432, p = .019) and NEIPS (B = 1.13, β = .373, p = .046) exhibited significant positive slopes, confirming that even in the early phase of the study, greater increases in emotional density and sentence-level intensity from II to DI were linearly related to larger gains in art interest. The other indicators, FEC, TEC, NFE, and TNEW, were not statistically significant ($p > .05$), suggesting that categorical diversity or emotional vocabulary breadth did not yet produce measurable linear effects on interest.

In week 2 (Table 5.10), the regression model revealed clearer linear trends. NFE (β = .383, p = .040), TNEW (β = .571, p = .001), NED (β = .580, p = .001), and NEIPS (β = .513, p = .004) all showed significant positive linear coefficients. Several constants in this week were negative (B = -1.03, p = .009), implying that, without emotional change, art interest tended to decline between sessions.

In week 3 (Table 5.10), the linear association between emotional change and art interest reached its strongest level. NED recorded the largest standardized coefficient across all models (β = .794, p < .001), representing the steepest positive slope linking emotional density to interest change. TNEW (β = .749, p < .001) and NEIPS (β = .653, p < .001) also showed strong linear effects, while TEC became significant (β = .417, p = .024). These results indicated that by the third week, participants who produced more emotionally compact, lexically diverse, and intense reflections exhibited proportionally greater gains in art interest. In contrast, FEC and NFE remained nonsignificant, suggesting that the linear trend in interest was driven primarily by textual rather than facial emotional development.

In week 4 (Table 5.10), the predictive structure stabilized, with NED (β = .528, p = .003) and NEIPS (β = .690, p < .001) maintaining their significance across all weeks. TNEW continued to show a moderate positive effect (β = .475, p = .009), while FEC became significant for the first time (β = .394, p = .034), indicating a minor late-stage contribution of facial emotional variety. TEC and NFE remained nonsignificant.

Several noteworthy values were observed across the four weeks (Table 5.10). The largest t-value appeared for NED in week 3 (t = 6.776, p < .001), confirming it as the most powerful linear predictor of interest improvement. NEIPS showed the second-strongest slope in week 4 (β = .690, t = 4.953, p < .001), demonstrating that the intensity of emotional expression continued to grow in predictive strength as the study progressed. NED and NEIPS were the only two indicators that remained significant throughout all four weeks, underscoring their stable linear association with changes in art interest. By contrast, FEC and TEC showed sporadic significance with smaller slopes, indicating that category diversity contributed only marginally to the prediction of interest gains.

Finally, NED and NEIPS again emerged as the most stable and significant predictors, displaying consistent positive linear relationships with changes in art liking (Table 5.11). The standardized coefficients of NED increased from 0.443 in week 1 to a maximum of 0.634 in week 2 and remained strong through week 4 (β = .616, all p < .001). Similarly, NEIPS maintained continuous significance across all four weeks, with β values ranging from 0.577 in week 1 to 0.665 in week 4 (p between .001 and < .001).

Table 5.11: Results of difference-score regression analysis between emotional indicators and art liking.

Predictors$^{\text{week 1}}$	B	SE	Beta	t	p
(Constant)	0.647	0.222		2.912	.007
FEC	0.124	0.255	0.094	0.488	.629
(Constant)	0.399	0.274		1.458	.156
TEC	0.280	0.186	0.278	1.506	.144
(Constant)	0.762	0.210		3.630	.001
NFE	-6.296	7.901	-0.152	-0.797	.432
(Constant)	0.268	0.360		.744	.463
TNEW	0.074	0.053	0.261	1.405	.171
(Constant)	0.204	0.261		0.785	.439
NED	29.991	11.667	0.443	2.517	.016*
(Constant)	0.023	0.244		0.092	.927
NEIPS	1.551	0.423	0.577	3.669	.001**
Predictors$^{\text{week 2}}$					
(Constant)	-0.560	0.316		-1.772	.088
FEC	0.520	0.292	0.324	1.782	.086
(Constant)	-0.049	0.323		-0.153	.880
TEC	-0.106	0.317	-0.064	-0.333	.742
(Constant)	-0.409	0.263		-1.555	.132
NFE	2.580	1.523	0.310	1.694	.102
(Constant)	-1.030	0.304		-3.390	.002
TNEW	0.235	0.065	0.574	3.641	.001**
(Constant)	-0.669	0.209		-3.198	.004
NED	60.134	14.105	0.634	4.263	$<$ .001***
(Constant)	-0.546	0.261		-2.093	.046
NEIPS	1.930	0.800	0.421	2.413	.023*
Predictors$^{\text{week 3}}$					
(Constant)	-0.196	0.386		-0.507	.616
FEC	0.083	0.283	0.057	0.295	.770
(Constant)	-0.177	0.309		-0.573	.571
TEC	0.074	0.209	0.067	0.351	.728
(Constant)	-0.154	0.275		-0.559	.581
NFE	0.526	1.626	0.062	0.324	.749
(Constant)	-0.985	0.311		-3.164	.004
TNEW	0.210	0.059	0.564	3.549	.001**
(Constant)	-0.654	0.243		-2.690	.012

Table 5.11 (continued)

NED	59.383	16.700	0.565	3.556	.001**
(Constant)	-0.494	0.231		-2.141	.041
NEIPS	1.909	0.607	0.518	3.146	.004**
Predictors$^{\text{week 4}}$					
(Constant)	0.021	0.377		0.056	.956
FEC	0.275	0.213	0.241	1.292	.207
(Constant)	0.052	0.281		0.184	.855
TEC	0.319	0.176	0.329	1.813	.081
(Constant)	0.166	0.254		0.652	.520
NFE	2.061	1.316	0.289	1.556	.129
(Constant)	-0.725	0.497		-1.461	.156
TNEW	0.095	0.038	0.435	2.513	.018*
(Constant)	-0.416	0.259		-1.608	.120
NED	32.823	8.074	0.616	4.065	< .001***
(Constant)	-0.391	0.229		-1.705	.100
NEIPS	1.369	0.296	0.665	4.622	< .001***

Note. FEC = facial emotion categories; TEC = textual emotion categories; NFE = non-dominant facial emotions; TNEW = total non-dominant emotion words; NED = non-dominant emotion density; NEIPS = non-dominant emotion intensity per sentence.
The dependent variable was the difference in mean art liking scores between the Delayed Impression (DI) and Initial Impression (II) sessions. The independent variables were the difference scores of six emotional indicators, calculated as DI minus II for each: FEC, TEC, NFE, TNEW, NED, and NEIPS.
* $p < .05$.; ** $p < .01$.; *** $p < .001$.

In week 1 (Table 5.11), both NED (B = 29.99, β = .443, p = .016) and NEIPS (B = 1.55, β = .577, p = .001) exhibited significant positive linear slopes, indicating that participants who showed greater increases in emotional density and intensity from II to DI tended to experience proportionally larger increases in liking toward the TBMA works. The other emotional indicators, FEC, TEC, NFE, and TNEW, were nonsignificant ($p > .05$).

In week 2 (Table 5.11), the linear effects of emotional predictors strengthened further. NED produced the strongest standardized coefficient across all weeks (β = .634, $p < .001$), representing a steep positive slope linking emotional density to liking change. TNEW also became significant (β = .574, p = .001), suggesting that greater emotional vocabulary usage corresponded to a proportional rise in liking. NEIPS remained significant (β = .421, p = .023), reinforcing its consistent predictive role. Notably, the constants in the NED and TNEW models were negative (B = -0.67 and -1.03, $p < .01$), indicating that without emotional elaboration, liking tended to decline between II and DI.

In week 3 (Table 5.11), NED ($\beta = .565$, $p = .001$) and NEIPS ($\beta = .518$, $p = .004$) continued to show significant positive linear effects, confirming the stability of emotional structure and intensity as predictors. TNEW also remained significant ($\beta = .564$, $p = .001$), marking the third consecutive week in which lexical richness contributed linearly to liking gains. Moreover, FEC, TEC, and NFE were again nonsignificant, while the constants were mostly negative and significant.

In week 4 (Table 5.11), NED ($\beta = .616$, $p < .001$) and NEIPS ($\beta = .665$, $p < .001$) remained highly significant, showing strong linear associations with liking improvement. TNEW maintained a moderate yet significant positive effect ($\beta = .435$, $p = .018$), while FEC, TEC, and NFE remained nonsignificant.

Several noteworthy patterns were evident across the models (Table 5.11). The strongest single predictive effect was observed for NED in week 2 (t = 4.263, $p < .001$), representing the steepest linear slope between emotional density and liking change. The second-highest effect occurred for NEIPS in week 4 (t = 4.622, $p < .001$), suggesting that sentence-level emotional intensity continued to accumulate predictive strength as participants engaged repeatedly with the artworks. NED and NEIPS were the only two indicators that achieved statistical significance in all four weeks, underscoring their temporal stability as linear predictors of liking.

5.4 Emotional and Temporal Predictors of Art Understanding, Interest, and Liking

To further examine whether emotional indicators predicted students' understanding of TBMA differently between the II and DI sessions, a series of Linear Mixed Models (LMM) were conducted in SPSS across the four-week observation period (Shek and Ma, 2011). Each model included TIME (II vs. DI) as a categorical fixed factor and one emotional indicator (FEC, TEC, NFE, TNEW, NED, or NEIPS) as a continuous covariate (Table 5.12). The variable DI was set as the reference level of TIME, so only the contrast term II appears in the output table. Therefore, the coefficient of II represents the mean difference in understanding scores between the II and DI sessions when the emotional indicator equals its grand mean (centered value). The Intercept corresponds to the estimated mean understanding score at DI when the emotional covariate equals zero. The main effect of each emotional indicator reflects its overall relationship with understanding across both sessions, while the interaction term (II $\times$ Emotion indicator) captures whether this emotional effect differs between II and DI conditions.

Across all four weeks (Table 5.12), the estimated II effect was negative, indicating that, within each week, expected understanding at DI exceeded that at II. This provided convergent evidence that re-viewing and delayed reflection yielded higher understanding than the initial impression for each of the four distinct TBMA works. NFE, NED, TNEW, and NEIPS, showed more informative associations with understanding than categorical diversity indices FEC and TEC.

Table 5.12: Estimated fixed-effects of time and emotional indicators on mean art understanding scores from linear mixed models.

Fixed Effects$^{\text{week 1}}$	B	SE	t	p
(Intercept)	5.376	0.198	27.094	< .001
II	-0.964	0.283	-3.410	.002**
FEC	0.722	0.470	1.537	.130
II × FEC	-1.018	0.530	-1.921	.060
(Intercept)	5.399	0.197	27.376	< .001
II	-0.972	0.277	-3.511	.001**
TEC	0.237	0.178	1.331	.189
II × TEC	-0.324	0.265	-1.223	.228
(Intercept)	5.499	0.157	34.968	< .001
II	-1.025	0.198	-5.174	< .001
NFE	5.276	1.370	3.852	< .001***
II × NFE	-5.248	1.778	-2.951	.006**
(Intercept)	5.330	0.208	25.613	< .001
II	-0.890	0.326	-2.728	.009**
TNEW	0.067	0.041	1.660	.103
II × TNEW	-0.079	0.076	-1.041	.304
(Intercept)	5.396	0.182	29.691	< .001
II	-0.857	0.248	-3.453	.001**
NED	15.771	7.966	1.980	.053
II × NED	-7.861	11.596	-0.678	.503
(Intercept)	5.478	0.191	28.692	< .001
II	-1.003	0.273	-3.681	.001**
NEIPS	0.218	0.346	0.631	.531
II × NEIPS	-0.214	0.588	-0.364	.718
Fixed Effects$^{\text{week 2}}$				
(Intercept)	5.313	0.246	21.619	< .001
II	-0.748	0.313	-2.391	.022*

Table 5.12 (continued)

FEC	0.014	0.391	0.036	.971
II × FEC	-0.127	0.510	-0.250	.804
(Intercept)	5.366	0.204	26.330	$< .001$
II	-0.780	0.261	-2.989	.005**
TEC	-0.125	0.226	-0.554	.582
II × TEC	0.057	0.290	0.196	.846
(Intercept)	4.864	0.219	22.171	$< .001$
II	-0.236	0.243	-0.970	.339
NFE	8.394	2.453	3.422	.001**
II × NFE	-8.107	2.781	-2.916	.006**
(Intercept)	5.248	0.206	25.481	$< .001$
II	-0.585	0.264	-2.214	.032*
TNEW	0.037	0.047	0.793	.431
II × TNEW	-0.011	0.061	-0.176	.862
(Intercept)	5.226	0.192	27.256	$< .001$
II	-0.512	0.229	-2.232	.033*
NED	20.667	14.158	1.460	.150
II × NED	2.003	17.207	0.116	.908
(Intercept)	5.234	0.190	27.540	$< .001$
II	-0.520	0.236	-2.208	.034*
NEIPS	0.784	0.548	1.431	.158
II × NEIPS	0.162	0.733	0.220	.827
Fixed Effects$^{\text{week 3}}$				
(Intercept)	5.352	0.169	31.616	$< .001$
II	-0.503	0.191	-2.633	.012*
FEC	-0.702	0.223	-3.144	.003**
II × FEC	0.803	0.245	3.285	.002**
(Intercept)	4.868	0.133	36.697	$< .001$
II	0.004	0.168	0.240	.981
TEC	0.195	0.114	1.703	.095
II × TEC	-0.037	0.194	-0.191	.850
(Intercept)	5.056	0.125	40.353	$< .001$
II	-0.227	0.145	-1.560	.0.129
NFE	-1.906	0.707	-2.695	.010*
II × NFE	2.672	1.511	1.769	.086
(Intercept)	4.809	0.135	35.738	$< .001$
II	0.113	0.175	0.647	.521
TNEW	0.074	0.030	2.475	.017*

Table 5.12 (continued)

II × TNEW	-0.012	0.047	-0.257	.799
(Intercept)	4.864	0.132	36.923	< .001
II	0.060	0.133	0.452	.654
NED	21.876	9.784	2.236	.031*
II × NED	6.388	11.837	0.540	.594
(Intercept)	4.874	0.123	39.506	< .001
II	0.016	0.132	0.119	.906
NEIPS	0.894	0.350	2.552	.014*
II × NEIPS	0.052	0.492	0.106	.916
Fixed Effects$^{\text{week 4}}$				
(Intercept)	5.233	0.209	24.941	< .001
II	-0.879	0.284	-3.096	.003**
FEC	-0.033	0.195	-0.172	.864
II × FEC	-0.299	0.279	-1.074	.289
(Intercept)	5.202	0.204	25.497	< .001
II	-0.573	0.240	-2.387	.023*
TEC	0.008	0.222	0.036	.971
II × TEC	0.018	0.303	0.060	.953
(Intercept)	5.093	0.164	31.009	< .001
II	-0.159	0.197	-0.805	.427
NFE	1.657	1.176	-2.695	.166
II × NFE	3.046	2.237	1.362	.180
(Intercept)	5.047	0.232	21.780	< .001
II	0.283	0.421	0.672	.505
TNEW	0.026	0.028	0.897	.375
II × TNEW	0.090	0.057	1.585	.122
(Intercept)	5.155	0.185	27.862	< .001
II	-0.162	0.221	-0.734	.468
NED	3.956	8.618	0.459	.649
II × NED	24.912	12.596	1.978	.055
(Intercept)	5.255	0.185	28.388	< .001
II	-0.334	0.214	-1.558	.128
NEIPS	-0.159	0.373	-0.426	.673
II × NEIPS	1.168	0.520	2.244	.030*

Note. II = Initial Impression; DI = Delayed Impression; FEC = facial emotion categories; TEC = textual emotion categories; NFE = non-dominant facial emotions; TNEW = total non-dominant emotion words; NED = non-dominant emotion density; NEIPS = non-dominant emotion intensity per sentence.

The dependent variable in all models is the mean score of art understanding. Each linear mixed

model (LMM) includes TIME (Initial Impression, II) as a categorical factor and one emotional indicator (FEC, TEC, NFE, TEW, NED, or NEIPS) as a continuous covariate. The second session (Delayed Impression, DI) was set as the reference level of TIME; therefore, only the contrast for II is displayed in the table. The intercept represents the estimated mean understanding score at the reference level (DI) when the emotional indicator is zero (mean-centered). The fixed effect of TIME (II) reflects the mean difference in understanding between II and DI. The fixed effect of each emotional indicator indicates how strongly that emotion predicts understanding across both sessions. The interaction term (II × Emotion) captures whether the relationship between the emotional indicator and art understanding differs between II and DI.
* $p < .05$.; ** $p < .01$.; *** $p < .001$.

From week 1 and week 2 (Table 5.12), NFE displayed positive estimated main effects at DI together with negative II × NFE interactions (week 1: Emotion B = 5.276, $p < .001$; II × Emotion B = -5.248, $p = .006$; week 2: Emotion B = 8.394, $p = .001$; II × Emotion B = -8.107, $p = .006$). Interpreted within each week, the slope at DI was sizable and positive, whereas the slope at II (main + interaction) was near zero. In week 3, NED, TNEW, and NEIPS showed positive estimated main effects without significant interactions (NED: B = 21.876, $p = .031$; TNEW: B = 0.174, $p = .017$; NEIPS: B = 0.894, $p = .014$). These results indicated similar positive slopes at DI and at II. NFE exhibited a sign switch (main = -1.906, $p = .010$; II × NFE = +2.672, $p = .010$): the slope at DI was negative, and the slope at II was slightly positive. In week 4, most main effects were not significant, but NEIPS presented a positive II × Emotion interaction (B = 1.168, $p = .030$) with a nonsignificant main effect, implying a stronger slope at II than at DI. NED displayed a marginal II × Emotion trend (B = 24.912, $p = .055$), hinting at a similar pattern.

Considering the four weeks as replications with different TBMA works (Table 5.12), three regularities stood out within each week and recurred across cycles. First, a clear time advantage was observed in every week, with higher understanding scores at DI than at II, confirming the stability of reflective gains across all TBMA works. Second, during week 1 and week 2, NFE demonstrated a DI-specific effect, showing strong positive slopes at DI but near-zero slopes at II (II × NFE: from -5.248 to -8.107, both $p = .006$), indicating that emotional reappraisal through facial expressivity primarily contributed to post-reflection understanding. Third, in the later weeks, the emotional predictors shifted toward text-based indicators, NED, TNEW, and NEIPS, which exhibited session-invariant or II-weighted slopes.

As for art interest, the fixed effects of time (II vs. DI) revealed a clear pattern of temporal adjustment in art interest (Table 5.13). In week 1, TIME coefficients were consistently negative (ranging from B = -0.547 to -0.476) and two out of six models reached significance ($p = .029$ and $p = .032$), demonstrating that participants' reported interest was significantly higher at

DI than at II. In week 2, all time estimates turned small and nonsignificant (ranging from B = -0.198 to 0.393, $p > .10$), indicating that the initial DI advantage largely disappeared, and participants' engagement stabilized between the two sessions. In week 3, the time coefficients remained modest and mostly nonsignificant, yet a clear modality divergence emerged between facial and textual indicators. FEC, TEC and NFE showed no significant effects, whereas all text-based non-dominant emotion indicators, TNEW, NED, and NEIPS, were highly significant. In week 4, TIME effects were small and inconsistent (ranging from B = -0.162 to 0.646), with only scattered significance (NED: B = 0.646, $p = .006$), reflecting no advantage of DI over II.

Regarding the Intercepts, which represent the estimated mean art-interest scores at DI when emotional indicators were mean-centered, values remained consistently high and stable across all weeks (Table 5.13). The intercepts ranged narrowly from approximately 4.4 to 4.9 in week 1, around 2.8 to 3.1 in week 2, 4.0 to 4.7 in week 3, and 4.1 to 4.5 in week 4, all highly significant ($p < .001$). These consistently positive and significant intercepts indicate that, regardless of individual emotional variations, participants maintained a robust baseline of interest toward TBMA works throughout the study. The transient dip in week 2 may reflect a momentary adjustment as participants encountered more complex or less familiar works, after which baseline interest rebounded and stabilized.

Across the four weeks and all six emotional models per week, in week 1 (Table 5.13), TEC had a significant positive main effect (B = 0.418, $p = .007$), indicating that greater textual emotion variety corresponded to higher overall art interest. Two text-based indicators, TNEW (B = 0.119, $p = .001$) and NED (B = 25.959, $p < .001$), were also strong positive predictors, revealing that both lexical richness and emotional compactness were linearly associated with heightened interest across sessions. Additionally, NEIPS (sentence-level emotional intensity) showed a positive effect (B = 0.886, $p = .005$), confirming that emotionally expressive writing enhanced art interest. No significant interaction terms were observed, implying stable effects across II and DI within this week.

Table 5.13: Estimated fixed-effects of time and emotional indicators on mean art interest scores from linear mixed models.

Fixed Effects$^{\text{week 1}}$	B	SE	t	p
(Intercept)	4.898	0.198	24.754	< .001
II	-0.547	0.239	-2.289	.029*

Table 5.13 (continued)

FEC	-0.548	0.461	-1.188	.240
II × FEC	0.697	0.509	1.369	.177
(Intercept)	4.561	0.171	26.716	$< .001$
II	0.178	0.271	0.655	.517
TEC	0.418	0.148	2.816	.007**
II × TEC	0.367	0.238	1.541	.129
(Intercept)	4.781	0.171	27.976	$< .001$
II	-0.476	0.211	-2.256	.032*
NFE	0.601	1.488	0.404	.688
II × NFE	-3.268	1.894	-1.725	.096
(Intercept)	4.441	0.180	24.628	$< .001$
II	0.426	0.290	1.465	.149
TNEW	0.119	0.035	3.371	.001**
II × TNEW	0.070	0.069	1.022	.312
(Intercept)	4.571	0.144	31.761	$< .001$
II	0.024	0.212	0.115	.910
NED	25.959	6.315	4.110	$< .001$***
II × NED	7.580	9.899	0.766	.449
(Intercept)	4.592	0.169	27.205	$< .001$
II	-0.016	0.239	-0.068	.946
NEIPS	0.886	0.306	2.898	.005**
II × NEIPS	0.291	0.514	0.566	.575
Fixed Effects$^{\text{week 2}}$				
(Intercept)	3.034	0.235	12.923	$< .001$
II	-0.198	0.327	-0.606	.548
FEC	-0.020	0.379	-0.054	.957
II × FEC	-0.313	0.497	-0.628	.532
(Intercept)	3.061	0.183	16.695	$< .001$
II	0.117	0.266	0.438	.664
TEC	-0.093	0.206	-0.455	.651
II × TEC	0.631	0.296	2.132	.039*
(Intercept)	2.810	0.226	12.419	$< .001$
II	0.196	0.289	0.678	.502
NFE	3.971	2.671	1.487	.143
II × NFE	-3.377	3.038	-1.111	.271
(Intercept)	2.937	0.166	17.643	$< .001$
II	0.393	0.238	1.650	.107
TNEW	0.046	0.039	1.205	.233

Table 5.13 (continued)

II × TNEW	0.139	0.057	2.451	.020*
(Intercept)	2.895	0.149	19.413	< .001
II	0.301	0.226	1.334	.192
NED	29.184	11.268	2.590	.012*
II × NED	20.365	16.060	1.268	.211
(Intercept)	2.912	0.149	19.578	< .001
II	0.309	0.229	1.353	.186
NEIPS	1.054	0.433	2.433	.018*
II × NEIPS	1.245	0.693	1.796	.080
Fixed Effects$^{\text{week 3}}$				
(Intercept)	4.708	0.259	18.126	< .001
II	-0.156	0.301	-0.517	.608
FEC	-0.378	0.349	-1.084	.284
II × FEC	0.739	0.387	1.918	.062
(Intercept)	4.279	0.181	23.590	< .001
II	0.288	0.239	1.206	.236
TEC	0.441	0.160	2.754	.008**
II × TEC	-0.012	0.273	-0.045	.964
(Intercept)	4.539	0.185	24.564	< .001
II	-0.245	0.245	-0.999	.325
NFE	-0.829	1.135	-0.730	.469
II × NFE	-0.415	2.508	-0.166	.869
(Intercept)	4.077	0.155	26.371	< .001
II	0.772	0.199	3.861	< .001***
TNEW	0.201	0.034	5.884	< .001***
II × TNEW	0.035	0.053	0.648	.521
(Intercept)	4.159	0.127	32.782	< .001
II	0.485	0.149	3.264	.003**
NED	73.419	10.364	7.084	< .001***
II × NED	-10.612	13.255	-0.801	.429
(Intercept)	4.276	0.142	30.203	< .001
II	0.303	0.175	1.732	.093
NEIPS	2.193	0.436	5.035	< .001***
II × NEIPS	0.007	0.635	0.011	.991
Fixed Effects$^{\text{week 4}}$				
(Intercept)	4.544	0.205	22.143	< .001
II	0.046	0.279	0.164	.870
FEC	0.555	0.191	2.896	.006**

Table 5.13 (continued)

II × FEC	-0.572	0.275	-2.080	.044*
(Intercept)	5.041	0.195	25.802	$< .001$
II	-0.112	0.255	-0.437	.664
TEC	-0.108	0.222	-0.487	.628
II × TEC	0.634	0.298	2.129	.038*
(Intercept)	4.884	0.174	28.089	$< .001$
II	-0.162	0.249	-0.652	.519
NFE	1.316	1.371	0.960	.341
II × NFE	0.411	2.535	0.162	.872
(Intercept)	4.416	0.215	20.526	$< .001$
II	1.192	0.404	2.949	.005**
TNEW	0.090	0.028	3.216	.002**
II × TNEW	0.072	0.057	1.257	.215
(Intercept)	4.442	0.151	29.342	$< .001$
II	0.646	0.225	2.879	.006**
NED	40.342	7.815	5.162	$< .001$***
II × NED	-3.414	11.716	-0.291	.772
(Intercept)	4.507	0.153	29.383	$< .001$
II	0.510	0.198	2.572	.014*
NEIPS	1.525	0.332	4.600	$< .001$***
II × NEIPS	-0.175	0.463	0.378	.707

Note. II = Initial Impression; DI = Delayed Impression; FEC = facial emotion categories; TEC = textual emotion categories; NFE = non-dominant facial emotions; NED = non-dominant emotion density; TNEW = total non-dominant emotion words; NEIPS = non-dominant emotion intensity per sentence.
The dependent variable in all models is the mean score of art interest. Each linear mixed model (LMM) includes TIME (Initial Impression, II) as a categorical factor and one emotional indicator (FEC, TEC, NFE, TEW, NED, or NEIPS) as a continuous covariate. The second session (Delayed Impression, DI) was set as the reference level of TIME; therefore, only the contrast for II is displayed in the table. The intercept represents the estimated mean interest score at the reference level (DI) when the emotional indicator is zero (mean-centered). The fixed effect of TIME (II) reflects the mean difference in interest between II and DI. The fixed effect of each emotional indicator indicates how strongly that emotion predicts interest across both sessions. The interaction term (II × Emotion) captures whether the relationship between the emotional indicator and art interest differs between II and DI.
* $p < .05$.; ** $p < .01$.; *** $p < .001$.

In week 2 (Table 5.13), TNEW and TEC retained their positive tendencies, while NED (B = 29.184, p = .012) and NEIPS (B = 1.054, p = .018) remained significant positive predictors. Importantly, TNEW displayed a significant interaction with TIME (II × TNEW, B = 0.139, p = .020), suggesting that the association between lexical emotional richness and art interest was stronger at II than at DI in this week. Similarly, TEC showed a modest positive interaction (II × TEC, B = 0.631, p = .039), implying that textual categorical emotions began to influence interest earlier in the appreciation process.

In week 3 (Table 5.13), TNEW demonstrated a strong positive effect (B = 0.201, p < .001), and TEC retained significance (B = 0.441, p = .008), implying that both lexical and categorical emotional variety predicted higher art interest regardless of timing. Similarly, NED (B = 73.419, p < .001) and NEIPS (B = 2.193, p < .001) emerged as dominant predictors, indicating that emotional compactness and intensity became stable, session-invariant predictors of art interest.

In week 4 (Table 5.13), FEC facial emotion categories had a positive main effect (B = 0.555, p = .006) with a significant negative interaction (II × FEC, B = -0.572, p = .044), indicating that greater facial emotional diversity predicted stronger interest at DI than II. TEC also displayed a significant positive interaction (II × TEC, B = 0.634, p = .038), suggesting that text-based variety had a delayed yet persistent impact. More importantly, NED (B = 40.342, p < .001) and NEIPS (B = 1.525, p < .001) remained powerful positive predictors, emphasizing that emotional compactness and sentence-level expressivity consistently supported sustained interest through the final week.

Regarding the results of art liking. The intercepts were consistently positive and highly significant (p < .001 across all models), ranging approximately from 4.2 to 5.1 throughout the four weeks. This indicates that, regardless of time or emotional variability, participants generally maintained a high level of liking toward the TBMA works. However, in the second week, the intercept dropped notably to around 2.8 to 2.9, which was significantly lower than the overall average. This suggests that the artwork presented during that week elicited relatively weaker affective responses, possibly due to differences in its visual or conceptual accessibility. Aside from this dip, the intercepts remained stable across the remaining weeks.

Table 5.14: Estimated fixed-effects of time and emotional indicators on mean art liking scores from linear mixed models.

Fixed Effects$^{\text{week 1}}$	B	SE	t	p
(Intercept)	5.097	0.209	24.281	< .001
II	-0.767	0.222	-3.448	.002**
FEC	-0.347	0.456	-0.761	.450

Table 5.14 (continued)

II × FEC	0.362	0.498	0.728	.471
(Intercept)	4.763	0.189	25.090	< .001
II	-0.262	0.235	-1.117	.271
TEC	0.491	0.163	3.010	.004**
II × TEC	-0.166	0.228	-0.728	.471
(Intercept)	5.032	0.186	27.003	< .001
II	-0.712	0.196	-3.634	.001**
NFE	-1.258	1.616	-0.778	.440
II × NFE	-0.206	1.761	-0.117	.908
(Intercept)	4.673	0.194	24.041	< .001
II	0.112	0.293	0.382	.704
TNEW	0.122	0.037	3.270	.002**
II × TNEW	0.036	0.066	0.546	.588
(Intercept)	4.756	0.145	32.878	< .001
II	-0.135	0.194	-0.696	.491
NED	32.765	6.326	5.179	< .001***
II × NED	2.866	9.071	0.316	.754
(Intercept)	4.724	0.159	29.747	< .001
II	-0.060	0.199	-0.303	.763
NED	1.386	0.276	5.002	< .001***
II × NED	0.157	0.418	0.375	.710
Fixed Effects$^{\text{week 2}}$				
(Intercept)	2.839	0.219	12.944	< .001
II	0.251	0.284	0.885	.382
FEC	0.220	0.351	0.628	.533
II × FEC	-0.146	0.458	-0.319	.751
(Intercept)	2.904	0.171	17.020	< .001
II	0.343	0.235	1.460	.153
TEC	0.070	0.192	0.365	.716
II × TEC	0.424	0.263	1.613	.115
(Intercept)	2.958	0.213	13.878	< .001
II	0.084	0.268	0.312	.757
NFE	-0.499	2.511	-0.199	.843
II × NFE	0.162	2.855	0.057	.955
(Intercept)	2.707	0.143	18.884	< .001
II	0.676	0.203	3.338	.002**
TNEW	0.117	0.033	3.517	.001**
II × TNEW	0.052	0.048	1.084	.286

Table 5.14 (continued)

(Intercept)	2.736	0.123	22.204	< .001
II	0.544	0.178	3.064	.004**
NED	43.396	9.405	4.614	< .001***
II × NED	5.706	12.931	0.441	.661
(Intercept)	2.783	0.132	21.009	< .001
II	0.514	0.198	2.591	.014*
NEIPS	1.371	0.388	3.531	.001**
II × NEIPS	0.821	0.607	1.354	.184
Fixed Effects$^{\text{week 3}}$				
(Intercept)	4.770	0.276	17.290	< .001
II	-0.015	0.334	-0.046	.964
FEC	-0.303	0.381	-0.7978	.429
II × FEC	0.390	0.429	0.909	.368
(Intercept)	4.466	0.197	22.609	< .001
II	0.271	0.273	0.991	.329
TEC	0.276	0.178	1.545	.128
II × TEC	-0.217	0.307	-0.706	.483
(Intercept)	4.600	0.185	24.831	< .001
II	-0.043	0.266	-0.164	.870
NFE	0.067	1.173	0.057	.955
II × NFE	-3.227	2.658	-1.214	.231
(Intercept)	4.245	0.165	25.667	< .001
II	0.903	0.241	3.746	.001**
TNEW	0.171	0.039	4.371	< .001***
II × TNEW	0.039	0.067	0.584	.563
(Intercept)	4.279	0.140	30.535	< .001
II	0.673	0.203	3.312	.002**
NED	69.800	12.113	5.763	< .001***
II × NED	-16.856	17.298	-0.974	.335
(Intercept)	4.371	0.146	29.937	< .001
II	0.545	0.208	2.623	.013*
NEIPS	2.276	0.466	4.882	< .001***
II × NEIPS	-0.235	0.714	-0.330	.743
Fixed Effects$^{\text{week 4}}$				
(Intercept)	4.788	0.225	21.295	< .001
II	-0.199	0.306	-0.652	.518
FEC	0.273	0.209	1.301	.199
II × FEC	-0.226	0.302	-0.749	.458

Table 5.14 (continued)

(Intercept)	4.959	0.198	25.013	$< .001$
II	-0.019	0.249	-0.076	.940
TEC	0.065	0.222	0.294	.770
II × TEC	0.561	0.301	1.866	.068
(Intercept)	4.931	0.182	27.089	$< .001$
II	-0.228	0.248	-0.919	.365
NFE	1.008	1.405	0.718	.476
II × NFE	1.202	2.625	0.458	.649
(Intercept)	4.417	0.219	20.108	$< .001$
II	1.287	0.412	3.120	.003**
TNEW	0.094	0.028	3.292	.002**
II × TNEW	0.092	0.058	1.576	.122
(Intercept)	4.512	0.164	27.427	$< .001$
II	0.501	0.223	2.247	.030*
NED	37.062	8.352	4.437	$< .001$***
II × NED	-2.065	12.349	-0.167	.868
(Intercept)	4.561	0.165	27.606	$< .001$
II	0.411	0.206	1.990	.054
NEIPS	1.433	0.351	4..088	$< .001$***
II × NEIPS	-0.062	0.489	-0.127	.899

Note. II = Initial Impression; DI = Delayed Impression; FEC = facial emotion categories; TEC = textual emotion categories; NFE = non-dominant facial emotions; TNEW = total non-dominant emotion words; NED = non-dominant emotion density; NEIPS = non-dominant emotion intensity per sentence.

The dependent variable in all models is the mean score of art liking. Each linear mixed model (LMM) includes TIME (Initial Impression, II) as a categorical factor and one emotional indicator (FEC, TEC, NFE, TEW, NED, or NEIPS) as a continuous covariate. The second session (Delayed Impression, DI) was set as the reference level of TIME; therefore, only the contrast for II is displayed in the table. The intercept represents the estimated mean liking score at the reference level (DI) when the emotional indicator is zero (mean-centered). The fixed effect of TIME (II) reflects the mean difference in liking between II and DI. The fixed effect of each emotional indicator indicates how strongly that emotion predicts liking across both sessions. The interaction term (II × Emotion) captures whether the relationship between the emotional indicator and art interest differs between II and DI.

* $p < .05$.; ** $p < .01$.; *** $p < .001$.

The fixed effects of time exhibited a clear but gradually diminishing DI advantage in liking. In week 1, time coefficients were negative and significant in multiple models (B = -0.767, p = .002; B = -0.712, p = .001), demonstrating that liking ratings were notably higher at DI than at II. In week 2, time effects weakened and became partly positive but nonsignificant in most models (B from about 0.25 to 0.68), suggesting a transitional stabilization between immediate and delayed impressions. In week 3, several positive and significant time coefficients re-emerged, particularly in text-based models (B = 0.903, p = .001; B = 0.673, p = .002; B = 0.545, p = .013), indicating that liking began to strengthen already during II. In week 4, time effects became small and inconsistent, with only a few models showing modest significance (B = 1.287, p = .003), confirming temporal convergence between II and DI.

Emotional predictors showed changing but complementary roles across the four weeks. In the early stage (week 1 to week 2), TEC, TNEW and NED were consistently significant (TEC: B = 0.491, p = .004; TNEW: B = 0.122, p = .002; NED: B = 32.765, $p < .001$), suggesting that richer emotional vocabulary and greater emotional density were associated with stronger liking. FEC and NFE, however, yielded nonsignificant results in most cases, indicating that subjective verbalization rather than visible emotion predicted affective preference. In week 3, text-based indicators again dominated (TNEW: B = 0.171, $p < .001$; NED: B = 69.800, $p < .001$; NEIPS: B = 2.276, $p < .001$), highlighting that as participants' emotional articulation matured, linguistic intensity and semantic differentiation became stable correlates of liking across both II and DI. In week 4, this pattern persisted: TNEW (B = 0.094, p = .002) and NED (B = 37.062, $p < .001$) continued to predict liking, reinforcing that emotional richness in verbal reflection sustained aesthetic preference.

Chapter 6

Discussion and Conclusion

6.1 Student Perceptions, Conditional Acceptance, and Implications of Emotion Recognition in Higher Art Education (From Study 1)

This study aimed to investigate undergraduate students' perceptions, concerns, and conditional acceptance regarding the use of smart mobile devices to record facial expressions in educational settings. Study 1, which sampled 199 third-year students from Shenyang Normal University, provided foundational insights into the intersection of Emotion Recognition Technology (ERT), mobile learning, and ethical acceptability in higher education. This chapter interprets the key findings, and reflecting on their theoretical, pedagogical, and technological implications.

6.1.1 General Acceptance, Functional Intrusiveness, and Ethical Concerns

In this study, students expressed a moderate level of general acceptance of integrating smart mobile devices into education, particularly when such devices were used for learning enhancement, engagement tracking, or video documentation. This finding highlights the increasing normalization of mobile technologies in higher education as flexible, personalized, and context-aware learning tools (Crompton and Burke, 2020; Hoi, 2020).

General acceptance was highest for academic uses, slightly lower for classroom photo or video recording, and notably more conservative when facial expression data were involved. This gradient reflects a well-documented phenomenon in which acceptance of educational technologies is often influenced by perceived functional intrusiveness (Venkatesh et al., 2003). The stratified acceptance levels may also be due to participants perceiving facial expression analysis as less "useful," more complex, or risk-laden compared to standard academic uses of mobile technologies. The results further revealed strong ethical and fairness-related concerns, particularly surrounding issues of privacy, bias, and consent. A significant portion of participants reported discomfort with being recorded, and even more expressed concerns about the potential for misinterpretation of facial data in academic assessment. These concerns mirror broader skepticism in the literature on educational emotion recognition technologies. For example, many facial expression classifiers contain cultural and gender-related biases (Barrett et al., 2019), potentially exacerbating inequities in evaluation. In this study, students' concerns extended beyond personal discomfort to structural fairness, a distinction that aligns with Selwyn's argument that ethical concerns in educational technology must be understood not only as matters of individual anxiety but also as issues of social justice (Selwyn, 2020). When the purposes and data flows are transparent, students tend to be receptive to learning analytics, they express strong reservations toward "Fitbit-style" tracking that may be perceived as leading to surveillance or intrusive monitoring (Rubel and Jones, 2016). Structural modeling of privacy concerns in learning analytics similarly indicates that apprehensions regarding secondary data use significantly reduce both the perceived benefits of data-driven analytics and individuals' behavioral intentions to adopt such systems (Mutimukwe et al., 2021). Within higher education learning analytics contexts (Viberg et al., 2018), privacy and informed consent have been consistently identified as among the most persistent barriers to adoption, even when the underlying technological solutions appear promising.

Regression analyses examining general acceptance and ethical concerns indicated that both variables were significant predictors of students' willingness to accept facial expression recording. General acceptance was positively associated with the acceptance of emotion recognition technologies (ERT), whereas ethical concerns exerted a negative influence. These findings confirm the multidimensional nature of student attitudes, suggesting that enthusiasm for technology does not automatically translate into endorsement of all its applications, particularly those perceived as sensitive or potentially intrusive. This dual pathway reflects what may be characterized as a form of "acceptance with concern." Accordingly, the study demonstrates that technology integration in educational contexts is jointly influenced by enabling and constraining factors. In the context of emotion recognition, enabling factors include participants' familiarity with and routine use of relevant functionalities (students' prior use of mobile devices for learning), while constraining factors center on perceived threats to autonomy and privacy.

Comparative analyses by gender and academic year indicated that female participants reported slightly higher levels of privacy concern and lower acceptance than male participants; however, these differences did not reach statistical significance. In contrast, third-year students demonstrated higher overall acceptance and lower resistance than their peers in other academic years. Accordingly, in samples with a predominance of female participants, greater attention should be paid to ethical and privacy-related considerations. The higher level of acceptance observed among third-year students may reflect greater pedagogical maturity and academic confidence, rendering them more receptive to novel feedback systems and experimental instructional approaches.

These findings informed the privacy-by-design considerations and participant selection strategies adopted in Studies 2 and 3 for the implementation of emotion recognition technologies.

6.1.2 Design and Implementation for Classroom Emotion Recognition

In this study, the open-ended responses revealed three interrelated categories of concern: emotional intrusion and psychological pressure; privacy and surveillance; and fairness and misinterpretation.

First, some students expressed concern that continuous recording by cameras or algorithms introduced a sense of compulsion and reduced their capacity for self-forgetful immersion in TBMA works. Emotion recognition technologies were perceived as potentially shifting students into a mode of metacognitive monitoring (questioning how they appear or what emotions the system is detecting) rather than supporting reflective engagement, thereby undermining the core spontaneity of the aesthetic experience. These concerns resonate with broader critiques of affective computing in education, which caution that persistent emotional monitoring risks transforming learners into objects of measurement rather than subjects of meaning-making. Andrejevic and Selwyn, in their analysis of facial recognition technologies in schools (Andrejevic and Selwyn, 2020), argue that such systems can easily shift from supportive tools to disciplinary mechanisms, normalizing cultures of continuous observation. The present findings reflect similar apprehensions, even within the relatively subjective context of art education, underscoring the importance of carefully defining when, how, and for what purposes emotional data are collected.

Second, privacy concerns in this study extended beyond issues of anonymity to encompass function creep. Students were particularly uneasy about the possibility that facial videos or emotional profiles might be stored long term, shared with third parties, or repurposed beyond their original instructional intent.

Third, participants raised concerns about whether facial expression datasets could accurately capture the nuanced and culturally mediated emotional responses elicited by TBMA, and whether inaccurate emotion labels (confusing anger with concentration) might negatively influence instructors' perceptions of them. Although the present study did not directly test algorithmic fairness, participants' evaluative responses indirectly reflected such anxieties. This is particularly salient given that the study focused on emotional data rather than behavioral traces (clicks, logins, or grades). In this respect, the findings resonate with research on biometric and health-related technologies (Dhagarra et al., 2020), where trust and perceived privacy risks have been shown to exert a substantial influence on user acceptance.

The inclusion of a multiple-choice item (Q16) in the instrument, which asked under what conditions students would feel more comfortable with being recorded, yielded valuable insights. The most frequently selected conditions included the provision of clearly explained informed consent, the ability to opt out at any time, and the option to use one's own device. Notably, a substantial proportion of participants expressed interest in the use of beautification filters, suggesting that aesthetic or appearance-related adjustments may be as meaningful to users as structural safeguards and autonomy. Collectively, these findings offer actionable design implications for future emotion-aware systems, underscoring the need to integrate technological innovation with user control and robust ethical frameworks.

Although Study 1 was primarily exploratory in nature, its findings exerted a direct influence on subsequent phases of the research project, particularly in relation to the appreciation of Time-Based Media Art (TBMA). Given that TBMA inherently involves temporality, affective engagement, and viewer reflection, the application of emotion recognition technologies in this domain requires both methodological rigor and heightened ethical sensitivity. The quantitative results of Study 1 suggest that integrating emotion recognition into TBMA-oriented curricula should prioritize transparency, voluntariness, and the preservation of student agency.

6.2 Re-viewing the Same Artwork and Emotional Reappraisal (From Study 2)

This study investigated how guided reflective re-viewing within a time-structured design mediates emotional reappraisal in TBMA by comparing two course sessions scheduled three days apart. Specifically, guided reflective re-viewing increased the breadth and intensity of non-dominant emotions, while the dominant emotion maintained a high degree of consistency.

6.2.1 The Roles of Negative and Positive Appraisals in Dominant and Non-Dominant Emotional Dynamics

The consistency analysis in this study revealed an intriguing discrepancy: while the dominant facial expressions were concentrated in "neutral" and "sad", which are typically associated with negative affect, the dominant emotions in the self-reported texts were predominantly positive, with the participants describing their experiences as "good". A possible explanation for this phenomenon is that the selected TBMA projects, characterized by abstraction, a slow pace, and melancholic aesthetic qualities, may have elicited low-arousal negative facial expressions (sad, neutral) during the guided-prompt durations. However, during the reflective evaluation (self-reported texts), the participants may have recognized cognitive or aesthetic appreciation, leading them to self-report more positive emotions. Alternatively, the participants may have suppressed or masked their negative emotional expressions in self-reports to conform to perceived social norms. This pattern is consistent with findings from previous studies on emotion regulation strategies. Gross emphasized the costs of emotional suppression, particularly its tendency to reduce external expression while increasing internal physiological stress (Gross and Levenson, 1997; John and Gross, 2004). Goldin et al. further distinguished between cognitive reappraisal and behavioral suppression (Goldin et al., 2008). In contrast to these studies, the present study focused on the more naturalistic aesthetic experiences induced by TBMA without explicit emotion regulation instructions. The findings suggest that even in the absence of strong external pressure to suppress or reappraise emotions, the participants spontaneously shifted from initially subdued or negative facial expressions to more positive retrospective self-assessments.

Another important finding of this study is that while the participants' dominant emotional expressions (sad, neutral, good, and dislike) remained stable between the initial and delayed impressions, non-dominant emotional features, such as the number of emotional categories, emotional density, and emotional intensity, significantly increased after three days during the delayed impressions. This finding aligns with the conclusions summarized in the review by Trupp et al. (Trupp et al., 2025), which indicate that reflective and educational accessory engagements serve as positive factors influencing outcomes rather than mere viewing alone (Bell and Robbins, 2007; Drake et al., 2024; Fancourt et al., 2020). However, this study further supplements existing research by suggesting that emotional changes during artistic engagement cannot be fully explained by positive factors alone. Specifically, deeper aesthetic engagement may depend on the dynamic interplay between positive and non-positive emotions (good, neutral, or sad) (Bjork et al., 2011). From the perspective of the experimental design of this study, sustained engagement and reflection facilitate emotional integration rather than the immediate enhancement of emotion. This process relies on the coexistence of positive and negative emotions elicited by the stimuli, fostering a more authentic sense of self-awareness and psychological resilience development. The differences in non-dominant emotions observed in this study make it plausible that the negative aesthetic emotions experienced during the initial impression (sad, neutral, dislike) acted as emotional catalysts, stimulating deeper emotional exploration during the reflective process. Moreover, these negative aesthetic emotions did not diminish over time; instead, they appeared to stabilize the participants' fundamental aesthetic orientation toward the art-works. Given prior research indicating that negative aesthetic emotions, such as sadness, can enhance critical thinking and cognitive engagement, and that individuals tend to al-locate more cognitive resources to processing negative stimuli (Schaefer et al., 2013; Silvia, 2009; Silvia and Brown, 2007), the present findings empirically challenge the traditional view that only positive emotions foster emotional broadening. Instead, the results suggest that negative aesthetic emotions can serve as potent drivers of emotional complexity and cognitive elaboration over time.

Moreover, rather than being driven solely by either positive or negative appraisals, the emotional outcomes were also strongly influenced by viewing duration optimization. Previous studies have indicated that overly short or excessively long guided-prompt durations do not necessarily enhance emotional engagement or comprehension, as they may lead to cognitive fatigue, overprocessing, or emotional detachment (Brieber et al., 2020; Codispoti et al., 2009). Therefore, this study adopted the 17s optimal presentation duration proposed by Brieber et al. to display the TBMA works under the DI condition (Brieber et al., 2020). However, the participants in this study differed significantly from those of Brieber et al., as all participants had received systematic training in art-related knowledge within an art school laboratory setting. This distinction highlights a pedagogical implication in which educators can tailor the length and pacing of guided reflection according to learners' prior experience, allowing sufficient cognitive space for emotional reappraisal without inducing fatigue. In this way, reflective viewing time functions as a flexible pedagogical variable that can be adjusted to balance emotional immersion and critical distance in art appreciation.

The experimental results of the present study supplement the work of Contreras-Medina et al. (González-Zamar and Abad-Segura, 2021), whose analysis emphasized the spatial and disciplinary structures of art education, mapping how visual culture operates as a cultural and epistemological system. In contrast, the current study advances this discussion by introducing temporal and affective dimensions through the lens of TBMA. By examining the emotional dynamics that unfold during repeated viewing, this research re-defines the educational ecosystem as a temporal ecology, one in which learning unfolds through iterative emotional reappraisal rather than through the static transmission of knowledge. While the analysis by Contreras-Medina et al. centered on the circulation of knowledge across artistic disciplines (Douglas and Coessens, 2012; Igdalova and Chamberlain, 2023; Schlegel et al., 2015), the present study emphasizes that affective circulation is an equally essential component of the educational ecosystem. The observed stability of dominant emotions and the diversification of non-dominant emotions across viewing stages illustrate an adaptive balance between cognitive regulation and emotional openness, a micro-level mechanism that reflects the macro-level adaptability emphasized in the literature on visual art ecosystems. Accordingly, the integration of structured re-viewing within the TBMA pedagogy proposed here exemplifies how future models of art education may translate macro-ecological goals, such as adaptability, creativity, and emotional literacy, into micro-level pedagogical practices. In this sense, the present study not only aligns with the educational ecosystem framework but also deepens it by shifting the focus from knowledge structures in visual arts toward affectively mediated learning processes in TBMA.

6.2.2 The Mere Exposure Effect and Conscious Reappraisal

Early studies on the Mere Exposure Effect (MEE) typically employed repeated presentations of simple visual stimuli, such as geometric figures, symbols, or faces, under tightly controlled laboratory conditions (Zajonc, 1968; Zajonc et al., 1974). These studies emphasized subliminal or unconscious exposure to index automatic affective preferences and showed that preference formation follows a non-linear function of exposure frequency that is consistent with associative learning and a curvilinear affective response, yet still operates through implicit associative mechanisms. Subsequent work by Monahan et al. extended this paradigm to examine affective diffusion (Monahan et al., 2000), demonstrating that repeated exposure can generate diffuse positive affect, even in the absence of awareness, without engaging conscious or reflective emotional processes. Zajonc further argued that MEE rests on noncognitive affective systems, reinforcing the view that the effect functions primarily at an unconscious level (Zajonc, 2001). Taken together, MEE research has conceptualized emotional change as a passive associative mechanism in which repetition increases perceptual fluency and thereby produces positive affect with-out conscious involvement.

By contrast, the present study relocates exposure within a conscious and reflective viewing context. The participants deliberately re-viewed the same TBMA video across a three-day interval. Rather than testing familiarity-based liking with simple stimuli, the de-sign used complex content, together with temporal structuring and guided reflection, enabling emotional reappraisal through memory retention and cognitive engagement. Furthermore, this study departed from strictly controlled laboratory settings by embedding emotion measurement in a naturalistic educational classroom environment. Moreover, whereas MEE research highlights unconscious affective diffusion, the present study indicates a layered affective development in which dominant emotions remain stable while non-dominant emotions become richer and more differentiated through conscious reappraisal. This shift from unconscious priming to conscious affective reorganization illustrates how reflective re-viewing could transform mere exposure into a process of aesthetic and pedagogical emotional learning.

6.2.3 Empirical Elaboration of Aesthetic Model and Meta-emotion Theory

Pelowski and Akiba's five-stage model of aesthetic experience was incorporated as the theoretical framework in this study to structure the exploration of students' emotional responses to TBMA (Pelowski and Akiba, 2011). First, the selection of participants, third-year art students who had received formal training in visual analysis, ensured that individuals possessed established perceptual schemas, thereby facilitating the pre-encounter stage. By presenting videos that were visually complex and stylistically diverse, the experimental design actively elicited the stage of cognitive mastery, triggering emotional and physiological arousal through visual dissonance or unfamiliarity. The inclusion of abstract and multilayered content was essential for encouraging participants to confront discrepancies, which is a prerequisite for the reevaluation of perceptual schemas. The layout of the tiered classroom significantly reduced social influence while enhancing the immersive quality of the viewing experience, thereby supporting the stage of secondary control, where unresolved tension could arise internally without external disruption. Furthermore, the time interval between the initial and delayed sessions facilitated meta-cognitive assessment, allowing participants sufficient time to reflect on and reinter-pret their initial emotional responses. The still-frame presentation format employed in the delayed session enabled focused re-engagement with the content, fostering opportunities for schema change or aesthetic outcomes, such as transformation or disconfirmation. This integration further reinforced the model's emphasis on the interplay of physiological, emotional, and cognitive processes in influencing art perception, providing a more concrete representation of how self-image and prior experiences influence aesthetic evaluation and transformation. Building on and extending Bartsch et al.'s meta-emotion theory (Bartsch et al., 2008), the present study's experimental procedures and findings align with their theoretical framework, particularly the conception of meta-emotion as a dynamic, appraisal-based process that influences media selection and emotion regulation, functioning as a continuous mediator throughout emotional engagement. Bartsch and colleagues advanced a broader conceptual model in which meta-emotion is understood as both a trait and a state, influenced by individual dispositions and situational contexts. While their framework emphasizes traditional appraisal dimensions such as novelty, goal conduciveness, and norm compatibility, the present study empirically tested the role of smart mobile devices and algorithmic feedback in triggering emotional responses, suggesting a novel, technologically mediated layer of appraisal. Furthermore, the extension proposed here shows how emerging media settings can amplify or modify emotional processes by implementing structured timing and guided viewing of TBMA content and incorporating feedback mechanisms that influence students' affective experiences. At the same time, emotion itself serves as the medium through which this process is appraised, further reinforcing the theoretical premise that emotion functions as a fundamental medium in the construction and internalization of aesthetic experience (Dennis et al., 2016; Mohammad et al., 2016; Simons et al., 1999).

6.2.4 Emotion Recognition as a Pedagogical Tool in Art Education

In this study, emotion recognition technology not only provided a quantitative means of assessing aesthetic responses but also functioned as an instructional aid in art class-rooms that was capable of sustaining attention and delivering timely feedback. Previous research has shown that emotion can modulate attention and learning (Kołakowska et al., 2014; Tonguç and Ozkara, 2020); therefore, visualizing affective states allows educators to adjust student engagement in real time. However, real-time modulation often requires higher hardware costs and more controlled classroom conditions (Nguyen et al., 2017). The auxiliary method proposed in this study offers a more practical approach for most art-teaching contexts while maintaining sufficient accuracy. Specifically, emotion measurements collected during the initial impression sessions can be aggregated to estimate the probability of core emotions over time and to flag small groups or the entire class when "attention decline" occurs (for instance, prolonged eye closure and absence of detectable facial expressions may indicate reduced attention). Educators can then incorporate a 60 s pause during the delayed impression session for guided reflection, pose targeted prompts, or replay a short segment of the TBMA work. Prior studies on vision-based attention monitoring and classroom analytics have confirmed the feasibility of such micro-interventions.

Within the context of reflective art viewing, technology-assisted emotion tracking may serve as a metacognitive tool that supports reflective learning (Bartsch et al., 2008; Mohammad et al., 2016). Rather than restricting emotion recognition to evaluative assessment, its integration into instructional design enables learners to externalize and monitor their emotional states, thereby fostering higher-order reflection and self-regulation. In addition to the objective detection of emotions from facial expressions, the self-reported texts provided by students offered rich introspective supplements that can inform pedagogical interventions in art education. Self-narratives often reveal emotional fluctuations tied to specific moments or themes within the artwork. For example, a student might note, "When the frame lingered on the decaying egg, I felt both sadness and curiosity," or "I hesitated here, feeling confused but also intrigued." Such affective cues allow educators to design micro-discussions or prompts that help students articulate latent meanings, connect emotional reactions to the formal or conceptual aspects of artworks, and transition from a sensory response to critical interpretation. Moreover, comparing students' self-reported descriptions with the emotions detected from their facial expressions may encourage reflection on moments of discrepancy, such as asking, "Why did the facial expressions detect surprise when I actually felt calm in my heart?" Through this continuous comparative process, learners can develop broader emotional vocabulary, stronger self-evaluation skills, and a more intentional interpretive stance. Over time, this iterative cycle of emotional attention, comparison, and verbalization may cultivate emotional literacy, deeper engagement, and more nuanced aesthetic judgment, contributing to the sustained innovation and development of art education.

6.3 On the Mutual Relation Between Emotion and Art Appreciation(From Study 3)

Study 3 sought to deepen the art appreciation of TBMA by extending the multimodal framework developed in Study 2. Through an expanded sample size, an open course setting, and weekly exposure to different TBMA works, this case study sought to examine how emotion recognition technologies, particularly facial expression and self-reported textual expression, could capture not only real-time emotional states but also reflective emotional changes. During four weeks, DI condition consistently showed higher levels of understanding than II condition, while interest and liking exhibited more fluctuating patterns. These findings extend study 1 and study 2 by demonstrating that the benefits of emotion-informed TBMA instruction generalize beyond single-session, single-work designs to a more authentic, longitudinal classroom context.

6.3.1 Replication of Emotional change Trends from Study 2

The present study extended the observation window to four weeks, presented a different TBMA work each week, and relocated viewing from a laboratory setting to a more open classroom environment. Under these new conditions, the emotional trajectory from II to DI reproduced the pattern observed in Study 2.

Specifically, negative feelings reported at II (sadness and disliking) appeared to lay the groundwork for subsequent change at DI. Previous research indicates that first unfavorable appraisal could leave internal cues that are reactivated upon re-encountering similar stimuli, thereby enabling emotional reappraisal; this phase is often marked by confusion or tension and is affectively negative (Nezlek and Kuppens, 2008; Parrott, 2002). In Study 3, the negative dominant emotions evident at II diminished by DI yet remained predominant, while shifting into more differentiated and nuanced forms, even though the TBMA materials and viewing context differed from those in Study 2. The persistence of this pattern suggests that the dominant negative response was not simply a by-product of a particular video genre or a constrained experimental setting; it also aligns with natural physiological responses that accompany focused viewing and reflective consideration (Monahan et al., 2000; Zajonc, 2001). At the same time, increases in non-dominant emotion (positive states: joy and good) cannot be attributed to mere repetition or the passage of time. Rather, structured spacing and guided viewing, which heightened attention and deliberate thinking, played a decisive role in influencing the observed changes.

Cognitive and behavioral inhibition likewise warrant attention (Barrett et al., 2001; Goldin et al., 2008; John and Gross, 2004). In the more open classroom setting of Study 3, the surrounding social context, even when it does not intrude on participants' personal space, can still influence display rules and self-presentation. This helps explain why textual reports skewed positive, whereas facial-expression analysis detected more negative affect. This dual-track pattern, which combines objective measurement with subjective reports and is embedded within the study's contextual scaffolding, captured externally oriented negative feelings expressed during viewing that were subsequently recoded linguistically into meaningful, value-laden self-reports. Very brief or very prolonged exposures can suppress engagement through underprocessing or fatigue (BJORK, 2011; Brieber et al., 2020). Given that participants in Study 3 had formal training in art, the same exposure durations may have supported deeper emotional reappraisal than would be expected for novices. Instructionally, pacing should be calibrated to validated findings on memory decay and optimal exposure time so that reflective prompts invite reinterpretation without inducing cognitive overload. Replicating the effect with a distinct corpus of stimuli and within a more ecologically valid classroom context therefore provides renewed empirical support for aesthetic models that emphasize the ongoing evolution of appraisal and affect (Leder and Nadal, 2014; Pelowski and Akiba, 2011; Trupp et al., 2025). Overall, the emotional trajectory observed in Study 3 should be interpreted not as a simple reproduction of prior results, but as a recalibration driven by emotional reappraisal.

6.3.2 Non-Dominant Emotions and Longitudinal Change in Understanding, Interest, and Liking (From Differential Regression)

In this study, difference of regression analyses indicate that the increase in emotion from DI to II can generally predict gains in art understanding, whereas its predictive power for increases in artistic interest and liking is comparatively weak. For art understanding, non-dominant facial expressions (NFE) and non-dominant emotion density (NED)from text emerged as the most stable predictors, showing statistically significant linear relationships across the four week period. For artistic interest and liking, NED and non-dominant emotion intensity per sentence (NEIPS) were the most stable predictors, likewise exhibiting statistically significant linear relationships across the same four week period.

Collectively, from II to DI, increases in emotion provide a more comprehensive account of the affective trace underlying gains in art understanding because such traces are evidenced in both facial and textual modalities. By contrast, the affective trace of interest and liking is manifested predominantly in self reported text, with limited support from facial impressions. Prior research commonly conceptualizes the development of interest as a progression from triggered situational interest to maintained situational interest and ultimately to a more stable individual interest (Specker et al., 2020). A comparable trajectory is often assumed for liking. Accordingly, unless additional conditions are met such as personal relevance, perceived value, autonomy, or identity congruence, short term emotional fluctuations may be insufficient to generate substantial increases in interest or liking. In other words, learners may become more capable of explaining TBMA while still lacking sustained motivational attachment to its content. Under such circumstances, cognitive gains can occur without commensurate gains in preference (Trope and Liberman, 2010).

This pattern is also consistent with the possibility that, when participants encounter the same work again after a time interval, they engage in an adjustment process that integrates initially ambiguous affect into more structured, reportable content. Notably, the stable predictors NFE, NED, and NEIPS are all non-dominant emotion indicators. Relative to dominant emotions, non-dominant emotions may be more likely to serve as catalysts for art appreciation. In the present study, emotion is conceptualized as a functional resource for attention, engagement, and memorability. The findings further suggest that temporal spacing enables receivers to express non-dominant emotions in meaningful ways within an art context. More specifically, facial expressions captured immediate disruptions or moments of imbalance during TBMA exposure, whereas textual data recorded nuanced, multicomponent affect articulated through reflective writing such as ambivalence or mixed emotions. For example, facial expression measures can effectively capture in the moment affective reactions during viewing such as surprise, tension, or confusion, whereas text based reports are better suited to capturing reflective appraisal, including what the emotion signifies and how it is interpreted. Consequently, art understanding appears to benefit from two complementary recognition pathways. One pathway is a direct in situ impact indexed by facial expressions. The other pathway is a reflective meaning making impact indexed by text. Together, these pathways enable multi layered prediction and may improve explanatory precision. By comparison, interest and liking, when measured exclusively through text, may be more susceptible to post hoc rationalization, linguistic norms, and perceived classroom expectations. Purely textual evaluations may therefore reflect socially influenced appraisal to a greater extent than raw affective response, which could attenuate the coupling between changes in emotion and changes in preference. Importantly, negative emotions may reflect attentional focus or cognitive strain (Cepeda et al., 2006; Friston, 2010; Scherer, 2009). Such negatively valenced, effortful states may be misinterpreted by participants as the absence of interest or liking, further weakening the observed association between emotional change and preference change.

6.3.3 Time and Emotion Effects on Understanding, Interest, and Liking (From LMM)

This study used linear mixed-effects models (LMM) to derive a more fine-grained account of how time, emotion, and art appreciation are interrelated. First, the main effect of time indicated that understanding at DI was consistently higher than at II, whereas liking showed smaller, stimulus-contingent time differences, and interest typically fell between these two outcomes. Second, indices capturing the structure and intensity of non-dominant emotions exhibited robust positive main effects across models, particularly for art understanding. Third, mixed effect (Time and Emotion) suggested that a given level of emotional signal was not systematically more consequential at DI than at II. This implies that re-encountering the same artwork influenced appreciation primarily through changes in emotional increment, and that emotion strength at either time point did not reliably predict higher or lower levels of appreciation.

The most stable result in the LMM analyses was that DI understanding exceeded II understanding with statistical significance. Time exerted a larger and more reliable influence on understanding, while interest and liking demonstrated smaller time-related differences that depended on stimulus characteristics. This pattern suggests that longitudinal practice with temporal spacing is robust across contexts. However, in most educational research, outcomes are typically operationalized as academic achievement, memory, or procedural performance rather than interpretive understanding of ambiguous aesthetic materials. For example, systematic syntheses in applied educational settings generally conclude that distributed practice supports achievement outcomes, while also emphasizing heterogeneity in both intervention formats and assessment types (Trumble et al., 2024). Extending this line of work, the present study indicates that, in the appreciation of TBMA, temporal spacing tends to support meaning-centered interpretation rather than hedonic endorsement. This finding aligns with the view that delayed re-viewing can shift learners from surface evaluation toward integrative interpretation. In other words, time did not merely increase familiarity; it also increased the likelihood that participants would construct more coherent descriptions of the work's meaning.

A second influential finding from the LMM analyses was the identification of emotion-related indicators that most strongly predicted appreciation outcomes. Across all models, indices of non-explicit emotional structure and intensity displayed strong positive main effects, especially for understanding. Relative to approaches that treat emotion as dominant-category labels or coarse positive–negative dimensions, the present modeling suggests that interpretive learning in TBMA is better quantified by how emotions are organized through time-spaced, guided re-encounters with the same artwork, as well as by the intensity with which they are expressed. This is consistent with constructionist and emotion-granularity perspectives, which propose that more context-sensitive and differentiated affective representations support adaptive cognition and regulation. Empirically, higher emotional granularity is associated with richer differentiation in everyday experience, suggesting that emotional specificity may co-vary with how experience is structured and accessed (Hoemann et al., 2021, 2023). The present study adds an education-specific implication: in TBMA settings, proxies related to granularity and indices related to intensity not only describe experience but also predict interpretive outcomes.

The LMM results also provide convergent support from text-based measures. Sentence-level expression intensity and richer verbal elaboration appeared to reflect deeper explanatory processing. Learners did not merely report an affective state; they actively coordinated perceptual cues, interpretive hypotheses, and justifications into a communicable form. This pattern is consistent with recent work in natural language processing, which argues that fine-grained sentence-level signals can capture psychologically meaningful affective structure and operate beyond simple lexical counts (Vishnubhotla, 2024). Unlike typical NLP applications focused on classification performance, this study treated multiple text indicators as covariates of instructional outcomes, thereby linking affective linguistic structure to learning-relevant appreciation rather than to categorical prediction alone.

The mixed effect (Time and Emotion) further yielded an important implication. Across many indices and outcomes, the strength of emotion–appreciation associations at DI was not uniformly greater than at II. Theoretically, this is valuable because it constrains a common interpretive shortcut in educational research that equates greater emotional intensity with universally better learning. At the same time, when time-dependent differences did emerge, they were not indiscriminate. Instead, the interaction patterns were more consistent with selective reweighting of specific affective traces, especially those reflecting integrative, non-dominant affect. DI did not render every emotional signal more predictive; rather, it appeared to increase the predictive value of emotional signals that were already constructed in ways compatible with explanation, contrast, and justification. This selective pattern can be interpreted through theories of conscious access and recursive integration. Compared with broad claims that "time changes everything," the findings support a narrower proposition: delayed re-encounter may increase the accessibility and inferential utility of previously weaker or non-dominant affective evidence, but only when that evidence is subsequently organized into a reportable form. This also provides a principled account of why art understanding benefited more reliably than interest and liking. If DI primarily enhances the availability of affective information for meaning-making and argumentation, the most stable gains should appear in understanding. Interest and liking should increase only when reorganized evaluations align with preference formation, which is more sensitive to stimulus properties and individual differences. Compared with spacing accounts that emphasize strengthened memory over time (Latimier et al., 2021), the spacing-related benefits observed in TBMA are more appropriately conceptualized as interpretive consolidation.

The observed DI advantage for understanding is consistent with applied syntheses suggesting that time-spaced, guided viewing can enhance durable learning beyond laboratory recall tasks. Accordingly, the primary demand in the present context was not the retention of facts, but the resolution of ambiguity and the construction of meaning.

6.3.4 Mechanisms and Design Principles for TBMA Learning

In this study, delayed re-viewing should not be interpreted as a simple repetition of exposure. Rather, it is more likely to reactivate an interpretive process that remained incomplete during the II. The temporal gap weakens immediate familiarity while simultaneously reintroducing uncertainty, such that when learners re-encounter the work at DI, they are more likely to reveal interpretive gaps and prediction errors. These discrepancies can, in turn, trigger further reflection, comparison, and integrative reasoning. Accordingly, the primary function of DI is not to consolidate an existing impression, but to promote meaning construction through reappraisal and reorganization. This mechanism also clarifies why time exerted a larger and more stable effect on understanding, whereas its effects on interest and liking were weaker and more contingent on stimulus characteristics. Understanding, as a meaning-oriented outcome, benefits more directly from processes of model updating and self-correction (Friston, 2010). By contrast, liking is more closely tied to immediate processing fluency and hedonic experience. When a work remains ambiguous and continues to demand substantial cognitive effort, hedonic fluency may not increase reliably over time and may even decline at II when interpretive cues are insufficient. In other words, for TBMA works that are open-ended and uncertain by nature, gains in understanding can develop more robustly because they are driven by the necessity of re-interpretation, whereas changes in interest and liking are more likely to be jointly influenced by stimulus properties, individual preferences, and situational engagement.

Motivation of agency further support this interpretation. Self-determination theory suggests that learners are more likely to engage deeply when their needs for autonomy, competence, and relatedness are supported, rather than disengaging from the task (Ryan et al., 2022). In TBMA contexts, some works may undermine the need for competence at II because ambiguity is high and interpretive cues are indeterminate, making it difficult for learners to obtain immediate feedback that "I can understand this." Under such conditions, even if interpretive processing begins to emerge, liking may decrease in the short term or remain unstable due to frustration or uncertainty. Cognitive load theory offers a compatible explanation. Instructional effectiveness depends on matching support to learners' knowledge states and individual differences, and even when learning ultimately occurs, working-memory constraints can suppress immediate positive experience (Sweller et al., 2019, 1998). From this perspective, DI may allow the cognitive burden associated with initial uncertainty to be transformed over time into more structured understanding, whereas interest and liking still depend on whether learners experience interpretive challenge as a productive form of engagement. This account is broadly consistent with the observed pattern in the present study: understanding increased most robustly, interest showed intermediate behavior, and liking exhibited weaker, more stimulus-dependent time effects.

Methodologically, an important improvement in this study was the use of LMM for repeated-measures data. Compared with approaches that collapse responses across individuals or treat repeated observations as independent, LMM separates between-person baseline differences from within-person change across time while simultaneously accounting for item-level variation. This is particularly critical for TBMA, given the inherent heterogeneity of responses across participants and artworks. In line with best-practice recommendations for mixed models, the selection of random-effects structures should be transparent, and unnecessarily complex specifications should be avoided when not supported by the data, as they can reduce power or hinder convergence (Matuschek et al., 2017; Meteyard and Davies, 2020). In this study, prioritizing interpretable fixed effects for time, emotion covariates, and their interactions supported a theoretically meaningful parameterization and aligned with contemporary expectations for confirmatory and reproducible mixed-model reporting.

Building on these mechanisms and findings, the results can be synthesized into a set of mutually reinforcing design principles for TBMA curricula. The central implication is that the robust DI advantage for understanding, together with the longitudinally predictive association between emotional traces and appreciation outcomes, these results support a curriculum organized around iterative II and DI cycles. Delayed re-engagement provides learners with opportunities to reactivate, test, and reconstruct interpretive hypotheses. Across repeated encounters, learners can develop more differentiated emotion concepts and more stable interpretive schemas, thereby converting uncertainty into defensible meaning construction.

First, DI should be designed as an explicit thinking and reasoning episode rather than as a mere re-presentation of the TBMA work. Course structure should therefore impose task demands at DI that require interpretation, comparison, and evidence-based articulation. DI prompts should be organized around questions such as what has changed and what evidence supports your interpretation, thereby encouraging learners to integrate perceptual cues, interpretive hypotheses, and justificatory reasons into a communicable argumentative structure. The goal of DI is not repeated experience, but the re-problematisation and reorganization of prior interpretations, enabling learners to respond to ambiguity through more integrative narratives. Second, scaffolding should be temporally asymmetric and aligned with the different instructional aims of II and DI. Rather than offering intensive guidance at both time points, a more effective strategy is to provide minimal, timely support during II to reduce unnecessary extraneous load while preserving exploratory space, allowing learners to generate initial interpretations and make their uncertainty visible. At DI, more stringent and challenging interpretive prompts should be introduced, requiring comparison, integration, and justification so as to capitalize on the DI phase as a window in which integrative processing is most predictive of understanding. This approach is consistent with cognitive load theory's emphasis on calibrating support to learners' knowledge states and the temporal dynamics of instruction (Sweller, 2024), and it facilitates the transformation of the burden of initial uncertainty into structured understanding. Third, motivationally supportive design is needed to translate gains in understanding into sustained interest and preference formation. Because liking in this study exhibited weaker and more stimulus-dependent time effects, spacing alone is unlikely to stabilize preference. Curricula should incorporate autonomy-supportive choices, visible markers of progress, and structured peer dialogue to strengthen learners' sense of competence and meaning. These supports increase the likelihood that interpretive challenge is experienced as competence building rather than as frustration, thereby stabilizing interest and gradually fostering liking over longer timescales (Ryan et al., 2022).

Finally, integrating emotion-recognition tools with structured reflection provides an implementable pathway for operationalizing these principles and for cultivating meta-emotional literacy. By visualizing emotional trajectories and inviting learners to articulate how and why their emotions change across II and DI, educators can support the development of conceptual and procedural knowledge about emotion regulation and meaning construction in art-educational contexts. The critical design claim advanced here is not to use emotion recognition as an external evaluative instrument, but to embed it within evidence-based reflective tasks, so that learners treat emotional change as interpretable, discussable, and revisable learning evidence. In this way, low-load support at II, high-demand prompts at DI, the legitimation of non-dominant emotions, and trajectory-based reflective structures can form an instructional loop that preserves the openness of TBMA while systematically promoting growth in interpretive understanding and providing motivational and cognitive conditions for more stable development of interest and liking.

6.4 Limitations and Future Research

This dissertation conducted empirical investigations of emotion and art appreciation in classroom settings. Although the work yields several contributions, a number of limitations warrant careful consideration. First, the review of TBMA education spanned the United States, Japan, and China; however, due to practical constraints, comparable sample sizes from the United States and Japan were not obtained. Consequently, across all three studies, participants were drawn from a single cultural and institutional context and were heavily concentrated in art majors. This sampling frame constrains external validity and may embed discipline-specific habits of seeing, discussing, and writing about art. In addition, the measurement strategy relied primarily on categorical facial-expression classifiers and lexicon-based analysis of Chinese texts. While these methods provide tractable, comparable indicators, they are not designed to capture the graded and often ambivalent aesthetic emotions characteristic of TBMA (awe, absorption, confusion). Finally, although robust informed-consent and privacy procedures were implemented at the individual level, the studies did not systematically assess institutional readiness (teacher training, data governance, workload implications), an important precondition for sustainable educational practice.

Study 1 used a questionnaire to assess the feasibility and ethical acceptability of emotion-recognition technologies. The instrument demonstrated generally acceptable internal consistency; nevertheless, it remains in an early stage of development, and several items require refinement. Because responses were self-reported within classroom contexts, they may have been subject to social desirability pressures and deference to instructor authority, potentially inflating willingness to be recorded or to adopt new technologies. The cross-sectional design further limits causal inference. Although subgroup comparisons were informative, the analyses do not adjudicate directional claims (whether heightened privacy concerns reduce willingness, or whether positive prior recording experiences increase acceptance).

Study 2 examined emotional change from II to DI within the same week. This design addresses short-term reappraisal, but it also introduces two constraints. First, the interval between II and DI was set to three days, which precluded estimating shorter or longer lags. Second, the DI session used static frames and guided prompts rather than continuous playback. Although pedagogically meaningful, some observed differences may reflect shifts in attentional allocation and memory retrieval induced by guided still images, thereby displacing emotions from those elicited by the TBMA work's temporal flow. With respect to emotion recognition, video capture relied on students' own devices, so variation in camera quality, angle, and lighting may have introduced systematic error into the facial-expression models. The lexicon-based approach to textual emotion likewise struggles to detect subtle aesthetic states or context-dependent nuance in student writing. Finally, the two-time-point design did not incorporate classroom-level covariates, such as instructor discourse or peer discussion after the session, that may moderate trajectories of emotional change between viewings.

Study 3 extended classroom inquiry to four weekly sessions and linked emotional indicators to understanding, interest, and liking. Its limitations lie chiefly in the analytic strategy. Although the difference-regression approach provided a transparent covariation index, it remains correlational: increasing familiarity with the course may drive parallel increases in both emotion and outcomes. In addition, the linear mixed models treated participants as random intercepts but did not include artwork and week as random factors; variance at the stimulus level may therefore be under-modeled. Finally, to mitigate multicollinearity, the six emotional indicators were analyzed in separate models, an appropriately cautious decision that nonetheless weakens within-model comparisons of their relative contributions.

In summary, these limitations indicate that the present evidence base should be interpreted as strong for contextual argumentation rather than as supporting broad causal generalization. Future research should therefore pursue cross-site replication across institutions, disciplines, and grade levels, with attention to gender balance. In emotion measurement, a hybrid categorical and dimensional framework is recommended: alongside discrete categories, continuous ratings of valence, arousal, and dominance should be collected, and aesthetic-specific scales (awe, confusion, immersion) incorporated. In data analysis, models that enter multiple emotional indicators simultaneously while controlling collinearity would allow estimation of relative weights and yield more interpretable results. Finally, because the studies were conducted in live classrooms, subsequent work should move beyond static frames toward more immersive formats while preserving temporal flow, e.g., 3D movie presentation of TBMA works to enrich in-class experience. By coordinating sampling, measurement, design, modeling and by integrating 3D presentation technologies with principled experimental controls, future studies may be able to draw stronger inferences about how emotional dynamics influence understanding, interest, and liking in media-art education.

6.5 Contribution to Knowledge Science

This dissertation contributes to Knowledge Science by positioning emotion as both a driver and a trace of knowing in art learning. Across three studies, it has combined dual-modal affect measurement (facial expressions and reflective text), a a structured-timing and guided-viewing design (initial versus delayed impressions), and analysis strategies that connect emotional change to understanding, interest, and liking. Building on these results, the contributions below synthesize academic, educational, and practical implications.

This research advances aesthetic theory by treating aesthetic experience as a temporally organized, revisable process rather than a static evaluation. It formalizes how emotional change across structured re-viewing as a revision, and it differentiates epistemic from hedonic pathways. It contributes an operational classroom aesthetic experience that links affective dynamics to learning-relevant outcomes. This dissertation reframes time and emotion as mediums whose temporal sequencing and guided attention are constitutive of meaning, not mere delivery conditions. By integrating a structured-timing and guided-viewing design with dual-modal affect measures, it specifies when and how the medium's organization modulates cognition and feeling. Concurrently, this research extends appraisal-based theory of meta-emotion by demonstrating a technologically mediated layer of appraisal. This reframes emotions in TBMA from static evaluations to dynamic, time-coupled processes that interact with presentation format and institutional setting. Together, these extensions provide a concrete, testable account of how affective dynamics, temporal design, and guided viewing co-produce aesthetic understanding, offering a portable framework for future cross-disciplinary inquiries in Knowledge Science.

At the pedagogical level, this research applies structured timing and guided interval viewing to the learning of TBMA-related knowledge, aligning shifts in emotional trajectories with gains in understanding and interest. This approach moves evaluation from static labels to trajectory-based assessment, encouraging educators to treat emotional reappraisal as a deliberate component of instructional design. By linking emotion indicators with art-appreciation outcomes, the framework yields quantifiable analytics that support actionable feedback. In the classroom, educators can time reflective prompts to coincide with peaks in emotional diversification, adjust pacing when arousal decays too rapidly, and decide whether to extend exposure or change works when confusion remains productive versus when it becomes obstructive.

For practice, the dissertation specifies the feasibility conditions for affect-aware instruction: use of student-owned devices, brief capture windows, transparent consent with opt-out options, and feedback channels that return results to both learners and educators. The findings indicate that lightweight data pipelines can generate decision-relevant signals without specialized laboratories, thereby lowering barriers to adoption in ordinary institutions. Beyond classrooms, the same measurement-feedback loop is adaptable to museum and gallery contexts. Curators may use aggregated audience-response profiles to place interpretive labels where confusion concentrates, optimize dwell time and spatial flow, and differentiate programs for novice versus expert audiences. Institutions can incorporate these analytics into iterative design cycles, strengthening evidence-based curation while maintaining robust privacy and data-governance safeguards.

References

Abdullah, M. U. and Alkan, A. (2022). A comparative approach for facial expression recognition in higher education using hybrid-deep learning from students' facial images. *Traitement du Signal*, 39(6).

Aiken, L. S., West, S. G., and Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions*. sage.

Aksu, G., Güzeller, C. O., and Eser, M. T. (2019). The effect of the normalization method used in different sample sizes on the success of artificial neural network model. *International journal of assessment tools in education*, 6(2):170–192.

Al Fakih, B., Morei, A., and Salehi, M. (2020). Students impression about using mobile phones in classroom. *International Journal of Emerging Technologies in Learning (IJET)*, 15(22):230–243.

Altman, D. G. and Bland, J. M. (1995). Statistics notes: the normal distribution. *Bmj*, 310(6975):298.

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., and Farhan, L. (2021). Review of deep learning: concepts, cnn architectures, challenges, applications, future directions. *Journal of big Data*, 8:1–74.

Andrejevic, M. and Selwyn, N. (2020). Facial recognition technology in schools: Critical questions and concerns. *Learning, Media and Technology*, 45(2):115–128.

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, 59(4):390–412.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3):255–278.

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., and Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological science in the public interest*, 20(1):1–68.

Barrett, L. F., Gross, J., Christensen, T. C., and Benvenuto, M. (2001). Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition & Emotion*, 15(6):713–724.

Bartsch, A., Vorderer, P., Mangold, R., and Viehoff, R. (2008). Appraisal of emotions in media use: Toward a process model of meta-emotion and emotion regulation. *Media Psychology*, 11(1):7–27.

Bell, C. E. and Robbins, S. J. (2007). Effect of art production on negative mood: A randomized, controlled trial. *Art Therapy*, 24(2):71–75.

Ben, X., Ren, Y., Zhang, J., Wang, S.-J., Kpalma, K., Meng, W., and Liu, Y.-J. (2021). Video-based facial micro-expression analysis: A survey of datasets, features and algorithms. *IEEE transactions on pattern analysis and machine intelligence*, 44(9):5826–5846.

Ben Fredj, H., Bouguezzi, S., and Souani, C. (2021). Face recognition in unconstrained environment with cnn. *The Visual Computer*, 37(2):217–226.

Birnhack, M. and Perry-Hazan, L. (2020). School surveillance in context: High school students' perspectives on cctv, privacy, and security. *Youth & Society*, 52(7):1312–1330.

Bjork, E. L., Bjork, R. A., et al. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the real world: Essays illustrating fundamental contributions to society*, 2(59-68):56–64.

BJORK, L. R. A. (2011). On the symbiosis of remembering, forgetting, and learning. In *Successful remembering and successful forgetting*, pages 19–40. Psychology Press.

Blair, R. C. and Higgins, J. J. (1985). Comparison of the power of the paired samples t test to that of wilcoxon's signed-ranks test under various population shapes. *Psychological Bulletin*, 97(1):119.

Bosch, N., D'Mello, S. K., Baker, R. S., Ocumpaugh, J., Shute, V., Ventura, M., Wang, L., and Zhao, W. (2016). Detecting student emotions in computer-enabled classrooms. In *IJCAI*, volume 16, pages 4125–4129.

Brieber, D., Forster, M., and Leder, H. (2020). On the mutual relation between art experience and viewing time. *Psychology of Aesthetics, Creativity, and the Arts*, 14(2):197.

Calvo, R. A. and D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on affective computing*, 1(1):18–37.

Canal, F. Z., Müller, T. R., Matias, J. C., Scotton, G. G., de Sa Junior, A. R., Pozzebon, E., and Sobieranski, A. C. (2022). A survey on facial emotion recognition techniques: A state-of-the-art literature review. *Information Sciences*, 582:593–617.

Carroll, N. (1985). The specificity of media in the arts. *Journal of Aesthetic Education*, 19(4):5–20.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., and Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological bulletin*, 132(3):354.

Chen, Q. and Yan, Z. (2016). Does multitasking with mobile phones affect learning? a review. *Computers in Human behavior*, 54:34–42.

Chen, Z., Wu, Z., Zamfir, E., Zhang, K., Zhang, Y., Timofte, R., Yang, X., Yu, H., Wan, C., Hong, Y., et al. (2024). Ntire 2024 challenge on image super-resolution (x4): Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6108–6132.

Chierico, A. (2016). Medium specificity in post-media practice. *V! RUS*, (12).

Ciobanu, P. and Fernaeus, Y. (2024). Material deconstructions of time: Posthumanist interventions through media art. *ACM Transactions on Computer-Human Interaction*, 31(6):1–47.

Codispoti, M., Mazzetti, M., and Bradley, M. M. (2009). Unmasking emotion: Exposure duration and emotional engagement. *Psychophysiology*, 46(4):731–738.

Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.

Cowan, N. and AuBuchon, A. M. (2008). Short-term memory loss over time without retroactive stimulus interference. *Psychonomic Bulletin & Review*, 15:230–235.

Cowie, R. and Cornelius, R. R. (2003). Describing the emotional states that are expressed in speech. *Speech communication*, 40(1-2):5–32.

Crompton, H. and Burke, D. (2018). The use of mobile learning in higher education: A systematic review. *Computers & education*, 123:53–64.

Crompton, H. and Burke, D. (2020). Mobile learning and pedagogical opportunities: A configurative systematic review of prek-12 research using the samr framework. *Computers & Education*, 156:103945.

Dennis, M., Masthoff, J., and Mellish, C. (2016). Adapting progress feedback and emotional support to learner personality. *International Journal of Artificial Intelligence in Education*, 26(3):877–931.

Dhagarra, D., Goswami, M., and Kumar, G. (2020). Impact of trust and privacy concerns on technology acceptance in healthcare: an indian perspective. *International journal of medical informatics*, 141:104164.

D'mello, S. K. and Kory, J. (2015). A review and meta-analysis of multimodal affect detection systems. *ACM computing surveys (CSUR)*, 47(3):1–36.

Douglas, A. and Coessens, K. (2012). Experiential knowledge and improvisation: Variations on movement, motion, emotion. *Art, Design & Communication in Higher Education*, 10(2):179–198.

Drake, J. E., Eizayaga, M., and Wawrzynski, S. (2024). Making versus viewing art: Effects on affect, enjoyment, and flow. *Empirical Studies of the Arts*, 42(1):147–165.

D'Mello, S. and Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, 22(2):145–157.

Ekanayake, S. Y. and Wishart, J. (2014). Mobile phone images and video in science teaching and learning. *Learning, Media and Technology*, 39(2):229–249.

Ekman, P. and Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2):124.

Ekman, P., Friesen, W. V., O'sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., LeCompte, W. A., Pitcairn, T., Ricci-Bitti, P. E., et al. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of personality and social psychology*, 53(4):712.

Elleström, L. (2010). The modalities of media: A model for understanding intermedial relations. In *Media borders, multimodality and intermediality*, pages 11–48. Springer.

Fancourt, D., Garnett, C., and Müllensiefen, D. (2020). The relationship between demographics, behavioral and experiential engagement factors, and the use of artistic creative activities to regulate emotions. *Psychology of Aesthetics, Creativity, and the Arts.*

Franěk, M., Petružálek, J., and Šefara, D. (2022). Facial expressions and self-reported emotions when viewing nature images. *International journal of environmental research and public health*, 19(17):10588.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, 11(2):127–138.

Fritz, C. O., Morris, P. E., and Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of experimental psychology: General*, 141(1):2.

Goldin, P. R., McRae, K., Ramel, W., and Gross, J. J. (2008). The neural bases of emotion regulation: reappraisal and suppression of negative emotion. *Biological psychiatry*, 63(6):577–586.

González-Zamar, M.-D. and Abad-Segura, E. (2021). Visual arts in the university educational ecosystem: Analysis of schools of knowledge. *Education Sciences*, 11(4):184.

Grau, O. (2010). *MediaArtHistories.* MIT press.

Greenberg, B. S. (1966). Media use and believability: Some multiple correlates. *Journalism Quarterly*, 43(4):665–670.

Gross, J. J. and Levenson, R. W. (1997). Hiding feelings: the acute effects of inhibiting negative and positive emotion. *Journal of abnormal psychology*, 106(1):95.

Gupta, S., Kumar, P., and Tekchandani, R. K. (2023). Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models. *Multimedia Tools and Applications*, 82(8):11365–11394.

Hadiprakoso, R. B., Setiawan, H., et al. (2020). Face anti-spoofing using cnn classifier & face liveness detection. In *2020 3rd International Conference on Information and Communications Technology (ICOIACT)*, pages 143–147. IEEE.

Hair, J. F., Risher, J. J., Sarstedt, M., and Ringle, C. M. (2019). When to use and how to report the results of pls-sem. *European business review*, 31(1):2–24.

Hansen, M. B. and Hansen, M. B. N. (2004). *New philosophy for new media*. MIT press.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Hoemann, K., Nielson, C., Yuen, A., Gurera, J. W., Quigley, K. S., and Barrett, L. F. (2021). Expertise in emotion: A scoping review and unifying framework for individual differences in the mental representation of emotional experience. *Psychological bulletin*, 147(11):1159.

Hoemann, K., Wormwood, J. B., Barrett, L. F., and Quigley, K. S. (2023). Multimodal, idiographic ambulatory sensing will transform our understanding of emotion. *Affective Science*, 4(3):480–486.

Höfel, L. and Jacobsen, T. (2003). Temporal stability and consistency of aesthetic judgments of beauty of formal graphic patterns. *Perceptual and Motor Skills*, 96(1):30–32.

Hoi, V. N. (2020). Understanding higher education learners' acceptance and use of mobile devices for language learning: A rasch-based path modeling approach. *Computers & Education*, 146:103761.

Hölling, H. (2017). *Paik's Virtual Archive: Time, Change, and Materiality in Media Art*. Univ of California Press.

Hutto, C. and Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, volume 8, pages 216–225.

Igdalova, A. and Chamberlain, R. (2023). Slow looking at still art: The effect of manipulating audio context and image category on mood and engagement during an online slow looking exercise. *Psychology of Aesthetics, Creativity, and the Arts*.

John, O. P. and Gross, J. J. (2004). Healthy and unhealthy emotion regulation: Personality processes, individual differences, and life span development. *Journal of personality*, 72(6):1301–1334.

Judd, C. M., Westfall, J., and Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: a new and comprehensive solution to a pervasive but largely ignored problem. *Journal of personality and social psychology*, 103(1):54.

Judd, C. M., Westfall, J., and Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual review of psychology*, 68(1):601–625.

Kanjo, E., Al-Husain, L., and Chamberlain, A. (2015). Emotions in context: examining pervasive affective sensing systems, applications, and analyses. *Personal and Ubiquitous Computing*, 19:1197–1212.

Katirai, A. (2024). Ethical considerations in emotion recognition technologies: a review of the literature. *AI and Ethics*, 4(4):927–948.

Keltner, D. and Haidt, J. (2003). Approaching awe, a moral, spiritual, and aesthetic emotion. *Cognition and emotion*, 17(2):297–314.

Ko, B. C. (2018). A brief review of facial emotion recognition based on visual information. *sensors*, 18(2):401.

Kołakowska, A., Landowska, A., Szwoch, M., Szwoch, W., and Wrobel, M. R. (2014). Emotion recognition and its applications. In *Human-computer systems interaction: Backgrounds and applications 3*, pages 51–62. Springer.

Krauss, R. E. (1999). Reinventing the medium. *Critical inquiry*, 25(2):289–305.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.

Küçük, A., Ganguly, A., Hao, Y., and Panahi, I. M. (2019). Real-time convolutional neural network-based speech source localization on smartphone. *IEEE Access*, 7:169969–169978.

Lai, C. and Zheng, D. (2018). Self-directed use of mobile devices for language learning beyond the classroom. *ReCALL*, 30(3):299–318.

Laird, R. D. and Weems, C. F. (2011). The equivalence of regression models using difference scores and models using separate scores for each informant: implications for the study of informant discrepancies. *Psychological assessment*, 23(2):388.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, pages 159–174.

Latimier, A., Peyre, H., and Ramus, F. (2021). A meta-analytic review of the benefit of spacing out retrieval practice episodes on retention. *Educational Psychology Review*, 33(3):959–987.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford university press.

Laurenson, P. (2013). Emerging institutional models and notions of expertise for the conservation of time-based media works of art. *Technè. La science au service de l'histoire de l'art et de la préservation des biens culturels*, (37):36–42.

Leder, H. (2013). Next steps in neuroaesthetics: Which processes and processing stages to study? *Psychology of Aesthetics, Creativity, and the Arts*, 7(1):27.

Leder, H. and Nadal, M. (2014). Ten years of a model of aesthetic appreciation and aesthetic judgments: The aesthetic episode–developments and challenges in empirical aesthetics. *British journal of psychology*, 105(4):443–464.

Leek, J., Rojek, M., Dobińska, G., and Kosiorek, M. (2024). Navigating the power of time in classroom practices: teachers' and students' perspectives. *Educational Review*, pages 1–23.

Li, J., Xu, Y., Xiong, H., and Wang, Y. (2010). Chinese text emotion classification based on emotion dictionary. In *2010 IEEE 2nd symposium on web society*, pages 170–174. IEEE.

Li, L., Chen, C. P., Wang, L., Liang, K., and Bao, W. (2023). Exploring artificial intelligence in smart education: Real-time classroom behavior analysis with embedded devices. *Sustainability*, 15(10):7940.

Liu, B. (2022). *Sentiment analysis and opinion mining*. Springer Nature.

Liu, Q. and Wang, L. (2021). t-test and anova for data with ceiling and/or floor effects. *Behavior Research Methods*, 53(1):264–277.

Luhmann, N. and Roberts, D. (1987). The medium of art. *Thesis Eleven*, 18(1):101–113.

Manovich, L. (2001). Post-media aesthetics. *Transmedia frictions, the digital, the arts, and the humanities*, 416.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing type i error and power in linear mixed models. *Journal of memory and language*, 94:305–315.

McLuhan, M. (2017). The medium is the message. In *Communication theory*, pages 390–402. Routledge.

Meigh-Andrews, C. (2013). *A history of video art*. A&C Black.

Menninghaus, W., Schindler, I., Wagner, V., Wassiliwizky, E., Hanich, J., Jacobsen, T., and Koelsch, S. (2020). Aesthetic emotions are a key factor in aesthetic evaluation: Reply to skov and nadal (2020).

Meteyard, L. and Davies, R. A. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, 112:104092.

Meyer, L. B. (2024). *Emotion and meaning in music*. University of chicago Press.

Misuraca, M., Forciniti, A., Scepi, G., and Spano, M. (2020). Sentiment analysis for education with r: packages, methods and practical applications. *arXiv preprint arXiv:2005.12840*.

Mittermeier, B. and Benade, L. (2024). Do innovative approaches to time allocation and timetable organisation provide a compelling alternative to traditional models? *Educational Review*, 76(7):1807–1824.

Mohammad, S., Shutova, E., and Turney, P. (2016). Metaphor as a medium for emotion: An empirical study. In *Proceedings of the fifth joint conference on lexical and computational semantics*, pages 23–33.

Mohammad, S. M. (2012). From once upon a time to happily ever after: Tracking emotions in mail and books. *Decision Support Systems*, 53(4):730–741.

Mohammad, S. M. (2016). Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In *Emotion measurement*, pages 201–237. Elsevier.

Mohammad, S. M. (2021). Sentiment analysis: Automatically detecting valence, emotions, and other affectual states from text. In *Emotion measurement*, pages 323–379. Elsevier.

Monahan, J. L., Murphy, S. T., and Zajonc, R. B. (2000). Subliminal mere exposure: Specific, general, and diffuse effects. *Psychological Science*, 11(6):462–466.

Monkaresi, H., Bosch, N., Calvo, R. A., and D'Mello, S. K. (2016). Automated detection of engagement using video-based estimation of facial expressions and heart rate. *IEEE Transactions on Affective Computing*, 8(1):15–28.

Mutimukwe, C., Twizeyimana, J. D., and Viberg, O. (2021). Students' information privacy concerns in learning analytics: Towards a model development. *arXiv preprint arXiv:2109.00068*.

Nandwani, P. and Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social network analysis and mining*, 11(1):81.

Nezlek, J. B. and Kuppens, P. (2008). Regulating positive and negative emotions in daily life. *Journal of personality*, 76(3):561–580.

Nguyen, B. T., Trinh, M. H., Phan, T. V., and Nguyen, H. D. (2017). An efficient real-time emotion detection using camera and facial landmarks. In *2017 seventh international conference on information science and technology (ICIST)*, pages 251–255. IEEE.

Nikolai, J. R., Bennett, G., Marks, S., and Gilson, G. (2019). Active learning and teaching through digital technology and live performance:'choreographic thinking'as art practice in the tertiary sector. *International Journal of Art & Design Education*, 38(1):137–152.

Nunnally, J. (1994). Psychometric theory.

Osborne, H. (1979). Some theories of aesthetic judgment. *The journal of aesthetics and art criticism*, 38(2):135–144.

Ota, K., Dao, M. S., Mezaris, V., and Natale, F. G. D. (2017). Deep learning for mobile multimedia: A survey. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 13(3s):1–22.

Pabba, C. and Kumar, P. (2022). An intelligent system for monitoring students' engagement in large classroom teaching through facial expression recognition. *Expert Systems*, 39(1):e12839.

Palm, K., Asp, A., and Håkansta, C. (2024). Implementing digital technologies in the school setting–how does it relate to work environment? *Educational Review*, pages 1–19.

Pang, B., Lee, L., et al. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval*, 2(1–2):1–135.

Pantic, M. and Rothkrantz, L. J. M. (2000). Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on pattern analysis and machine intelligence*, 22(12):1424–1445.

Parrott, W. G. (2002). The functional utility of negative emotions. *The wisdom in feeling: Psychological processes in emotional intelligence*, pages 341–359.

Pelowski, M. and Akiba, F. (2011). A model of art perception, evaluation and emotion in transformative aesthetic experience. *New ideas in psychology*, 29(2):80–97.

Peppler, K. A. (2010). Media arts: Arts education for a digital age. *Teachers College Record*, 112(8):2118–2153.

Perignat, E. and Katz-Buonincontro, J. (2019). Steam in practice and research: An integrative literature review. *Thinking skills and creativity*, 31:31–43.

Phillips, J. (2012). Shifting equipment significance in time-based media art. *Electronic Media Review*, 1:139–154.

Picard, R. W. (2000). Toward computers that recognize and respond to user emotion. *IBM systems journal*, 39(3.4):705–719.

Picard, R. W. and Healey, J. (1997). Affective wearables. *Personal technologies*, 1:231–240.

Poria, S., Cambria, E., Bajpai, R., and Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information fusion*, 37:98–125.

Potts, A. (2004). Tactility: the interrogation of medium in art of the 1960s. *Art history*, 27(2):282–304.

Quaranta, D. (2013). *Beyond new media art*. Link Editions Brescia.

Rafique, H., Liu, M., Lin, Q., and Yang, T. (2022). Weakly-convex–concave min–max optimization: provable algorithms and applications in machine learning. *Optimization Methods and Software*, 37(3):1087–1121.

Rancière, J. (2011). What medium can mean. *Parrhesia*, 11:35–43.

Rappleye, J. and Komatsu, H. (2016). Living on borrowed time: Rethinking temporality, self, nihilism, and schooling. *Comparative education*, 52(2):177–201.

Ravaja, N., Kallinen, K., Saari, T., and Keltikangas-Jarvinen, L. (2004). Suboptimal exposure to facial expressions when viewing video messages from a small screen: Effects on emotion, attention, and memory. *Journal of Experimental Psychology: Applied*, 10(2):120.

Rinehart, R. and Ippolito, J. (2022). *Re-collection: Art, new media, and social memory*. MIT Press.

Rubel, A. and Jones, K. M. (2016). Student privacy in learning analytics: An information ethics perspective. *The information society*, 32(2):143–159.

Ryan, R. M., Duineveld, J. J., Di Domenico, S. I., Ryan, W. S., Steward, B. A., and Bradshaw, E. L. (2022). We know this much is (meta-analytically) true: A meta-review of meta-analytic findings evaluating self-determination theory. *Psychological Bulletin*, 148(11-12):813.

Saarni, C. (1999). *The development of emotional competence*. Guilford press.

Schaefer, S. M., Morozink Boylan, J., Van Reekum, C. M., Lapate, R. C., Norris, C. J., Ryff, C. D., and Davidson, R. J. (2013). Purpose in life predicts better emotional recovery from negative stimuli. *PloS one*, 8(11):e80329.

Scheerens, J., Hendriks, M. A., Luyten, J. W., Sleegers, P., and Glas, C. A. (2013). Productive time in education. a review of the effectiveness of teaching time at school, homework and extended time outside school hours.

Scherer, K. R. (2009). The dynamic architecture of emotion: Evidence for the component process model. *Cognition and emotion*, 23(7):1307–1351.

Schlegel, A., Alexander, P., Fogelson, S. V., Li, X., Lu, Z., Kohler, P. J., Riley, E., Tse, P. U., and Meng, M. (2015). The artist emerges: Visual art learning alters neural structure and function. *NeuroImage*, 105:440–451.

Schuller, B., Batliner, A., Steidl, S., and Seppi, D. (2011). Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. *Speech communication*, 53(9-10):1062–1087.

Selwyn, N. (2020). Re-imagining 'learning analytics'… a case for starting again? *The Internet and Higher Education*, 46:100745.

Semeniuta, S., Severyn, A., and Barth, E. (2016). Recurrent dropout without memory loss. *Preprint arXiv:1603.05118*.

Shaik, T., Tao, X., Dann, C., Xie, H., Li, Y., and Galligan, L. (2023). Sentiment analysis and opinion mining on educational data: A survey. *Natural Language Processing Journal*, 2:100003.

Shao, J. and Wu, D. (2024). Evaluation on algorithms and models for multi-modal information fusion and evaluation in new media art and film and television cultural creation. *Journal of Computational Methods in Science and Engineering*, 24(4-5):3173–3189.

Sharples, M., Taylor, J., and Vavoula, G. (2010). A theory of learning for the mobile age: Learning through conversation and exploration across contexts. *Medienbildung in neuen Kulturräumen: die deutschprachige und britische Diskussion*, pages 87–99.

Shek, D. T. and Ma, C. M. (2011). Longitudinal data analyses using linear mixed models in spss: concepts, procedures and illustrations. *The scientific world journal*, 11(1):42–76.

Sheng, H. and Lau, M. (2024). Optimising real-time facial expression recognition with resnet architectures. *Journal of Machine Intelligence and Data Science (JMIDS)*, 5(1):33–45.

Sherring, A. (2020). Divergent conservation: cultural sector opportunities and challenges relating to the development of time-based art conservation in australasia. *AICCM Bulletin*, 41(1):69–82.

Shrestha, N. (2021). Factor analysis as a tool for survey analysis. *American journal of Applied Mathematics and statistics*, 9(1):4–11.

Silvia, P. J. (2005). Emotional responses to art: From collation and arousal to cognition and emotion. *Review of general psychology*, 9(4):342–357.

Silvia, P. J. (2009). Looking past pleasure: anger, confusion, disgust, pride, surprise, and other unusual aesthetic emotions. *Psychology of Aesthetics, Creativity, and the Arts*, 3(1):48.

Silvia, P. J. and Brown, E. M. (2007). Anger, disgust, and the negative aesthetic emotions: Expanding an appraisal model of aesthetic experience. *Psychology of Aesthetics, Creativity, and the Arts*, 1(2):100.

Sim, J. and Wright, C. C. (2005). The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical therapy*, 85(3):257–268.

Simons, R. F., Detenber, B. H., Roedema, T. M., and Reiss, J. E. (1999). Emotion processing in three systems: The medium and the message. *Psychophysiology*, 36(5):619–627.

Singh, R., Saurav, S., Kumar, T., Saini, R., Vohra, A., and Singh, S. (2023). Facial expression recognition in videos using hybrid cnn & convlstm. *International Journal of Information Technology*, 15(4):1819–1830.

Sophonhiranrak, S. (2021). Features, barriers, and influencing factors of mobile learning in higher education: A systematic review. *Heliyon*, 7(4).

Specker, E., Forster, M., Brinkmann, H., Boddy, J., Pelowski, M., Rosenberg, R., and Leder, H. (2020). The vienna art interest and art knowledge questionnaire (vaiak): A unified and validated measure of art interest and art knowledge. *Psychology of Aesthetics, Creativity, and the Arts*, 14(2):172.

Spielmann, Y. (2012). *Hybrid culture: Japanese media arts in dialogue with the West*. Mit Press.

Sun, Y., Ni, C.-C., and Kang, Y.-Y. (2023). Comparison of four universities on both sides of the taiwan strait regarding the cognitive differences in the transition from stem to steam in design education. *Education Sciences*, 13(3):241.

Sung, Y.-T., Chang, K.-E., and Liu, T.-C. (2016). The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. *Computers & Education*, 94:252–275.

Swami, V. (2013). Context matters: Investigating the impact of contextual information on aesthetic appreciation of paintings by max ernst and pablo picasso. *Psychology of Aesthetics, Creativity, and the Arts*, 7(3):285.

Sweller, J. (2024). Cognitive load theory and individual differences. *Learning and Individual Differences*, 110:102423.

Sweller, J., Van Merriënboer, J. J., and Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational psychology review*, 31(2):261–292.

Sweller, J., Van Merrienboer, J. J., and Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational psychology review*, 10(3):251–296.

Tai, J. H.-M., Bellingham, R., Lang, J., and Dawson, P. (2019). Student perspectives of engagement in learning in contemporary and digital contexts. *Higher Education Research & Development*, 38(5):1075–1089.

Tan, E. S. (2013). *Emotion and the structure of narrative film: Film as an emotion machine*. Routledge.

Tanwar, R., Phukan, O. C., Singh, G., Pal, P. K., and Tiwari, S. (2024). Attention based hybrid deep learning model for wearable based stress recognition. *Engineering Applications of Artificial Intelligence*, 127:107391.

Toba, H., Hernita, Y. T., Ayub, M., and Wijanto, M. C. (2024). Bloom-epistemic and sentiment analysis hierarchical classification in course discussion forums. *arXiv preprint arXiv:2402.01716*.

Tonguç, G. and Ozkara, B. O. (2020). Automatic recognition of student emotions from facial expressions during a lecture. *Computers & Education*, 148:103797.

Trope, Y. and Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological review*, 117(2):440.

Trumble, E., Lodge, J., Mandrusiak, A., and Forbes, R. (2024). Systematic review of distributed practice and retrieval practice in health professions education. *Advances in Health Sciences Education*, 29(2):689–714.

Trupp, M. D., Howlin, C., Fekete, A., Kutsche, J., Fingerhut, J., and Pelowski, M. (2025). The impact of viewing art on well-being—a systematic review of the evidence base and suggested mechanisms. *The Journal of Positive Psychology*, pages 1–25.

Tse, M. M., Ng, J. K., Chung, J. W., and Wong, T. K. (2002). The effect of visual stimuli on pain threshold and tolerance. *Journal of Clinical Nursing (Wiley-Blackwell)*, 11(4).

van de Vall, R. (2022). Theories of time-based media art conservation: From ontologies to ecologies. In *Conservation of Time-Based Media Art*, pages 13–27. Routledge.

Van den Hoven, J., Lokhorst, G.-J., and Van de Poel, I. (2012). Engineering and the problem of moral overload. *Science and engineering ethics*, 18:143–155.

Vassos, S., Malliaraki, E., Falco, F. d., Di Maggio, J., Massimetti, M., Nocentini, M. G., and Testa, A. (2016). Art-bots: Toward chat-based conversational experiences in museums. In *Interactive Storytelling: 9th International Conference on Interactive Digital Storytelling, ICIDS 2016, Los Angeles, CA, USA, November 15–18, 2016, Proceedings 9*, pages 433–437. Springer.

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, pages 425–478.

Viberg, O., Hatakka, M., Bälter, O., and Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in human behavior*, 89:98–110.

Vinodhini, G. and Chandrasekaran, R. (2012). Sentiment analysis and opinion mining: a survey. *International Journal*, 2(6):282–292.

Vishnubhotla, K. (2024). *Computational Measures of Language Variation in Textual Utterances*. PhD thesis, University of Toronto (Canada).

Vistorte, A. O. R., Deroncele-Acosta, A., Ayala, J. L. M., Barrasa, A., López-Granero, C., and Martí-González, M. (2024). Integrating artificial intelligence to assess emotions in learning environments: a systematic literature review. *Frontiers in psychology*, 15:1387089.

Wallbott, H. G. and Scherer, K. R. (1989). Assessing emotion by questionnaire. In *The measurement of emotions*, pages 55–82. Elsevier.

Walton, K. L. (1993). How marvelous! toward a theory of aesthetic value. *The journal of aesthetics and art criticism*, 51(3):499–510.

Wang, J., Cao, B., Yu, P., Sun, L., Bao, W., and Zhu, X. (2018). Deep learning towards mobile applications. In *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*, pages 1385–1393. IEEE.

Wang, M., Shen, R., Novak, D., and Pan, X. (2009). The impact of mobile learning on students' learning behaviours and performance: Report from a large blended classroom. *British Journal of Educational Technology*, 40(4):673–695.

Westfall, J., Kenny, D. A., and Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, 143(5):2020.

Whitehill, J., Serpell, Z., Lin, Y.-C., Foster, A., and Movellan, J. R. (2014). The faces of engagement: Automatic recognition of student engagementfrom facial expressions. *IEEE Transactions on Affective Computing*, 5(1):86–98.

Williams, B., Onsman, A., and Brown, T. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian journal of paramedicine*, 8:1–13.

Yadollahi, A., Shahraki, A. G., and Zaiane, O. R. (2017). Current state of text sentiment analysis from opinion to emotion mining. *ACM Computing Surveys (CSUR)*, 50(2):1–33.

Yantaç, A. E., Özcan, O., and Emengen, A. (2011). Participatory design in interactive media design education for the solution of unfamiliar design problems: a case study on a disabling environment and an emerging technology. *Digital Creativity*, 22(1):40–48.

Yong, A. G., Pearce, S., et al. (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in quantitative methods for psychology*, 9(2):79–94.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of personality and social psychology*, 9(2p2):1.

Zajonc, R. B. (2001). Mere exposure: A gateway to the subliminal. *Current directions in psychological science*, 10(6):224–228.

Zajonc, R. B., Markus, H., and Wilson, W. R. (1974). Exposure effects and associative learning. *Journal of Experimental Social Psychology*, 10(3):248–263.

Zhao, X. (2025). Breakthroughs, lags, and contradictions: an analysis of the practices and effects of new media art aesthetic education in china. In *Frontiers in Education*, volume 10, page 1483559. Frontiers Media SA.

Zhou, H., Huang, M., Zhang, T., Zhu, X., and Liu, B. (2018). Emotional chatting machine: Emotional conversation generation with internal and external memory. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

Zou, K. H., Tuncali, K., and Silverman, S. G. (2003). Correlation and simple linear regression. *Radiology*, 227(3):617–628.