# Strengthening Design and Evidence in Digital Game-Based Learning: Aligning Comparisons, Isolating Design Features, and Integrating Outcomes and Research Agendas

Dissertation
zur Erlangung des Doktorgrades
der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Eberhard Karls
Universität Tübingen

vorgelegt von
Amedeo Viccari, M.Sc.

Tübingen
2026

| | |
|---|---|
| 1. Betreuer: | Prof. Dr. Richard Göllner |
| 2. Betreuer: | Prof. Dr. Jens-Uwe Hahn |
| 3. Betreuer: | Prof. Dr. Kou Murayama |
| | |
| Tag der mündlichen Prüfung: | 03.07.2026 |
| Dekanin: | Prof. Dr. Taiga Brahm |
| Dekan: | Prof. Dr. Dominik Papies |
| | |
| 1. Gutachter: | Prof. Dr. Richard Göllner |
| 2. Gutachter: | Prof. Dr. Kou Murayama |

# ACKNOWLEDGEMENTS

This dissertation would not have been possible without the support of a number of people whose guidance, trust, and expertise were essential throughout my doctoral journey.

First and foremost, I would like to thank Prof. Dr. Richard Göllner, my first supervisor, for his trust in my project and his continuous support throughout the PhD. His confidence in the overall direction of my work and his openness to my interdisciplinary research interests provided a stable and encouraging framework in which this dissertation could develop.

I am also very grateful to Prof. Dr. Jens-Uwe Hahn, my second supervisor, for his support and for giving me the opportunity to work and learn at the Hochschule der Medien in Stuttgart. My time there was particularly important for the development of this dissertation, as it allowed me to acquire essential skills in game development and to gain hands-on experience in designing and implementing my own educational video game. In this context, I also thank Prof. Dr. Sabiha Ghellal for her valuable advice, expertise, and constructive feedback during this phase. I also wish to express my appreciation to Prof. Dr. Kou Murayama for serving as my third supervisor and for his support, as well as for agreeing to act as my examiner. I would like to thank Prof. Dr. Luise von Keyserlingk for taking on the role of chair of my defence, and for her work on behavioural data in learning contexts, which has been highly inspiring to me.

I would also like to thank Dr. Anna Kienitz for her crucial support in setting up the experimental studies and for her rigorous methodological guidance during my first statistical analysis. Her precise feedback and high standards had a lasting impact on the quality of my work. I am equally grateful to Prof. Dr. Marlit Lindner for her thoughtful and honest feedback, which was particularly important during some of the most challenging phases of my research. A special thanks goes to Dr. Aki Schumacher for her support during the writing process of this dissertation. Her careful reading, constructive comments, and critical questions helped clarify arguments and improve the overall structure and quality of my work. I also thank the Hector Institute and LEAD Graduate School and Research Network for their support and for providing a stimulating research environment.

My deep gratitude goes to Prof. Dr. Lisa Bardach. Throughout the entire PhD process, she provided constant support, guidance, and intellectual encouragement. Her methodological rigor, critical perspective, and reliability profoundly shaped both the empirical and conceptual development of this dissertation. I am deeply grateful for the opportunity to learn from such an outstanding researcher, and I am convinced that this project would not have reached its current form without her continuous mentorship and support.

Last but not least, I would like to thank my parents and my siblings: Their trust and support were essential throughout this journey.

*Ringrazio mia madre, mio padre, mio fratello e mia sorella per il loro sostegno costante lungo tutto il mio percorso; il loro supporto ha fatto la differenza, permettendomi di affrontare ogni difficoltà con maggiore sicurezza e determinazione.*

*Inoltre, voglio dedicare questa tesi anche ai miei due nipotini, Filippo e Matteo: non abbiate mai paura di sbagliare; le migliori lezioni nella vita vengono dagli errori che commettiamo, ed è proprio questo che ci aiuta a crescere.*

# ABSTRACT

Digital game-based learning is increasingly used in educational contexts, yet evidence regarding its effects remains difficult to evaluate. A central challenge lies in recurring methodological limitations in the field, including weak comparison logic between game and non-game conditions, limited isolation of specific design features, and fragmentation in outcome constructs and research agendas. These issues complicate the interpretation and interpretability of findings and hinder the development of cumulative knowledge on the functioning of digital games.

To address these challenges, this dissertation developed two purposefully designed and structurally aligned learning environments, described in two design papers: *Limes*, an educational video game on Roman history, and a corresponding interactive text-based application containing identical narrative content, tasks, and informational resources. These systems enable controlled comparisons between instructional formats while holding key elements constant and allow for the systematic manipulation of specific design features.

Building on this foundation, two pre-registered experimental studies examined how format and design features influence cognitive, motivational, emotional, and behavioural engagement outcomes in digital history learning. The first study compared game-based and text-based learning under different instructional framings (learning vs. fun). The second experiment isolated a specific design feature by manipulating the presence of background audio within the game. Across both studies, behavioural log data were used to capture learners' interaction with optional informational resources.

Results showed that the game condition elicited higher levels of triggered situational interest and enjoyment than the text condition (small effects). There were no significant effects on maintained interest and boredom. In addition, playing the game did not lead to higher knowledge scores, as compared to the text condition. Instructional framing did not affect the investigated outcomes, and did not interact with format. With respect to the second experiment, it was shown that the manipulation of background audio did not have significant effects on the investigated cognitive, motivation, emotional and behavioural engagement outcomes. Across both studies, however, behavioural engagement with content-rich informational resources (Codex entries), as assessed using log data, was consistently associated with higher knowledge performance.

Complementing the design and experimental work and extending beyond the domain of history learning, a systematic review on video game effects on intercultural outcomes (prejudice and intercultural competence) was conducted. The review synthesized a heterogeneous body of research across different lines of inquiry to address fragmentation in outcome constructs and research agendas.

The review highlighted fragmentation in outcome constructs, particularly for intercultural competence, with limited conceptual clarity and a narrow focus on selected components such as empathy and perspective-taking, while knowledge and behavioural aspects of intercultural competence were rarely assessed. At the same time, it identified distinct research strands focusing on different mechanisms and game features (e.g., collaboration, embodiment, format comparisons), thereby providing a structured perspective that organizes existing studies by their research foci and the mechanisms they examine.

Overall, this dissertation's findings indicated that outcomes in digital game-based learning are often not determined by format categories or design features alone but seem to be closely related to how learning-relevant content is structured and how learners engage with that content. Methodologically, the dissertation demonstrated the value of structurally aligned comparison designs, feature-level experimental manipulation, and in-game behavioural measurement. By integrating system development, experimental investigation, and systematic synthesis, the dissertation contributed to a more precise understanding of how digital game-based learning environments can be designed and tested in educational research.

# ZUSAMMENFASSUNG

Digitales spielbasiertes Lernen (im Englischen: Digital Game-Based Learning) wird zunehmend in Bildungskontexten eingesetzt, dennoch bleibt die Evidenz zu seinen Effekten oftmals schwer zu bewerten. Eine zentrale Herausforderung besteht in wiederkehrenden methodischen Einschränkungen des Forschungsfeldes, darunter eine schwache Vergleichslogik zwischen spielbasierten und nicht-spielbasierten Bedingungen, eine unzureichende Isolation spezifischer Designmerkmale sowie eine Fragmentierung von Outcome-Konstrukten und Forschungsansätzen. Diese Probleme erschweren die Interpretation von Befunden und die Entwicklung einer kumulativen Evidenzbasis zur Funktionsweise von Videospielen als Lernumgebungen.

Um diesen Herausforderungen zu begegnen, entwickelte die vorliegende Dissertation zwei gezielt konzipierte und strukturell aufeinander abgestimmte Lernumgebungen, die in jeweils einem Artikel beschrieben werden: *Limes*, ein edukatives Videospiel zur römischen Geschichte, sowie eine dazugehörige interaktive textbasierte App mit identischen narrativen Inhalten, Aufgaben und Informationsressourcen. Diese Systeme ermöglichen kontrollierte Vergleiche zwischen Instruktionsformaten bei gleichzeitiger Konstanthaltung zentraler Elemente und eröffnen die Möglichkeit, spezifische Designmerkmale systematisch zu manipulieren.

Auf dieser Grundlage untersuchen zwei präregistrierte experimentelle Studien, wie Format und Designmerkmale kognitive, motivationale, emotionale sowie verhaltensbezogene Engagement-Outcomes im digitalen Geschichtslernen beeinflussen. Die erste Studie verglich spielbasiertes und textbasiertes Lernen unter unterschiedlichen instruktionalen Rahmungen (Instruktion mit dem Fokus auf Lernen vs. Spaß). Die zweite Studie isolierte ein spezifisches Designmerkmal durch die Manipulation von Hintergrundaudio im Spiel. In beiden Studien werden Logdaten zum Verhalten der Lernenden genutzt, um ihre Interaktion mit optionalen Informationsressourcen zu erfassen.

Die Ergebnisse des ersten Experiments zeigen, dass die Spielbedingung im Vergleich zur Textbedingung höhere Werte für ausgelöstes („triggered“) situationales Interesse und Freude hervorrief (Effekte mit kleiner Effektstärke). Für aufrechterhaltenes („maintained“) situationales Interesse und Langeweile zeigten sich keine signifikanten Unterschiede. Zudem führte die Spielbedingung nicht zu höheren Wissensleistungen im Vergleich zur Textbedingung.

Die instruktionale Rahmung wies weder Effekte auf die untersuchten Outcomes noch eine Interaktion mit dem Format auf. Im zweiten Experiment zeigte sich, dass die Manipulation von Hintergrundaudio keine signifikanten Effekte auf kognitive, motivationale, emotionale oder verhaltensbezogene Engagement-Outcomes hatte. Über beide Studien hinweg war jedoch das verhaltensbezogene Engagement mit inhaltsreichen Informationsressourcen (Codex-Einträgen), erfasst über Logdaten, konsistent mit höheren Wissensleistungen verbunden.

Ergänzend zu den Design- und Experimentalstudien und über den Bereich des Geschichtslernens hinaus wurde ein systematisches Review zu den Effekten von Videospielen auf interkulturelle Outcomes (Vorurteile und interkulturelle Kompetenz) durchgeführt. Das Review synthetisierte ein heterogenes Forschungsfeld und arbeitete die Fragmentierung von Outcome-Konstrukten und Forschungsagenden heraus.

Die Ergebnisse des Reviews zeigten eine Fragmentierung der Outcome-Konstrukte, insbesondere im Bereich der interkulturellen Kompetenz. Diese Fragmentierung äußerte sich in geringer konzeptueller Klarheit sowie in einer Fokussierung auf einzelne Komponenten wie Empathie und Perspektivübernahme, während Wissens- und Verhaltensaspekte interkultureller Kompetenz selten erfasst werden. Gleichzeitig identifizierte das Review unterschiedliche Forschungsstränge, die sich auf verschiedene Mechanismen und Spielmerkmale konzentrieren (z. B. Kooperation, Embodiment, Formatvergleiche). Es liefert damit einen strukturierenden Rahmen, um bestehende Studien nach ihren Forschungsschwerpunkten und untersuchten Mechanismen systematisch einzuordnen.

Insgesamt zeigen die Befunde dieser Dissertation, dass Outcomes im digitalen spielbasierten Lernen häufig nicht allein durch Formatkategorien oder einzelne Designmerkmale bestimmt werden, sondern eng damit zusammenhängen, wie lernrelevante Inhalte strukturiert sind und wie Lernende mit ihnen interagieren. Methodisch verdeutlicht die Arbeit den Erkenntnisgewinn durch strukturell vergleichbare Designs, die experimentelle Isolation einzelner Designmerkmale sowie der Erhebung von Verhaltensdaten innerhalb digitaler Lernumgebungen. Durch die Integration von Systementwicklung, experimenteller Untersuchung und systematischer Synthese leistet die Dissertation einen Beitrag zu einem präziseren Verständnis der Gestaltung und empirischen Untersuchung digitaler spielbasierter Lernumgebungen.

# TABLE OF CONTENTS

# 1

# INTRODUCTION AND THEORETICAL FRAMEWORK

# 1 Introduction and Theoretical Framework

Digital technologies have become a fundamental part of contemporary educational contexts, shaping how learners access information and engage with instructional content (e.g., OECD, 2015, 2021). Across formal and informal settings, digital environments increasingly influence the conditions under which learning takes place, offering new possibilities for interaction and engagement. Within this broader development, digital games have received increasing attention as learning environments in educational research (e.g., Clark et al., 2016; Ninaus et al., 2019; Plass et al., 2015; Siuko et al., 2025).

A central reason for this interest lies in the distinct characteristics of digital games compared to more traditional instructional formats. Digital games are characterized by interactive environments that involve goal-directed activity, feedback systems, and structured challenges, requiring learners to actively engage with the system. Many games also incorporate features such as narrative framing and progression structures (e.g., levels, quests) (e.g., Gee, 2007; Plass et al., 2015). Accordingly, they can be conceptualized as learning environments in which learners are not only exposed to information but also interact with structured systems that respond to their actions. These characteristics suggest that that digital games may influence how learners engage with instructional content and how learning processes unfold (E. A. Boyle et al., 2016; Clark et al., 2016; Plass et al., 2015; Wouters et al., 2013).

Research on digital game-based learning has therefore examined a broad range of academic outcomes (e.g., Plass et al., 2015). In addition to cognitive outcomes such as knowledge acquisition, studies have also examined how video games influence learners' emotional experiences, motivation, and engagement (e.g., Clark et al., 2016; Ninaus et al., 2019; Wouters et al., 2013). Together, these findings show that effects of digital games extend beyond knowledge acquisition to include multiple other dimensions of academic outcomes.

Moreover, while less commonly considered in educational research, digital games also hold potential to influence intercultural outcomes (e.g., Adachi et al., 2015; Bachen et al., 2012; Behm-Morawitz, Hoffswell, et al., 2016; Breves, 2020; Peña et al., 2018). In particular, it has been discussed to what extent the use of digital games and specific features implemented within these games can help to reduce prejudice and influence intercultural competences (e.g., Costa et al., 2024; Peng et al., 2010). For example, digital games may expose learners to diverse perspectives, narratives, and social situations, allowing them to engage with viewpoints that differ from their own experiences (e.g., Gutierrez et al., 2014).

Still, despite a growing body of research, the empirical evidence on the effects of digital games remains mixed, and some areas are still insufficiently understood (e.g., Clark et al., 2016; Lee & Chen, 2023; Wouters et al., 2013). One source of this variability lies in the design and comparison logic of existing studies. Comparisons between game-based and non-game formats often differ in multiple dimensions, such as interactivity, feedback, and pacing, making it difficult to attribute observed effects to the game format itself. In addition, digital games typically combine multiple design features, which are rarely isolated experimentally, making it difficult to determine which specific features contribute to observed effects. Finally, particularly in the context of intercultural outcomes, research is characterized by heterogeneous constructs and divergent research foci, limiting the accumulation of consistent evidence. Together, these factors complicate the interpretation and comparability of findings and impede a more coherent understanding of how digital games function both as learning environments for academic outcomes and as contexts for the development of intercultural outcomes.

The present dissertation aimed to address these limitations. Two purposefully designed learning environments—a digital game and a structurally aligned interactive text application for history learning—were developed, enabling controlled comparisons between formats and the systematic manipulation of design features. Building on this foundation, two experimental studies examined how format (game versus text) and design features (audio versus no audio within the game) influence cognitive, motivational, emotional, and behavioural engagement outcomes in digital history learning. This work is complemented by a systematic review on digital game effects on prejudice and intercultural competence, extending the scope beyond history learning. The review addresses fragmentation in outcome constructs and research agendas and identifies cross-domain conceptual and methodological challenges related to intercultural outcomes.

The dissertation is structured as follows. First, Chapter 1 establishes the theoretical and conceptual foundations of digital game-based learning. It introduces key definitions and distinctions (Chapter 1.1), reviews the role of digital games as learning environments and their effects on academic outcomes, with a particular focus on history learning (Chapter 1.2), and examines research on prejudice and intercultural competence (Chapter 1.3). It then outlines central methodological and conceptual challenges in DGBL research, including weak comparison logic, confounded design features, and fragmented outcome constructs and research agendas (Chapter 1.4), before positioning the scope and contribution of the present dissertation (Chapter 1.5). Building on this foundation, Chapter 2 presents the aims and research questions

of the dissertation. These are organized around three core objectives: strengthening comparison logic through structurally aligned formats and instructional framing, isolating specific design features within digital games, and clarifying outcome constructs and research agendas in the field. Chapter 3 provides the methodological foundation through the two design papers, which document the development of the educational video game *Limes* and the corresponding interactive text-based application. These systems form the basis for the empirical studies described next. Chapters 4 and 5 present the empirical studies. Chapter 4 examines the effects of instructional format (game vs. text) and instructional framing (learning vs. fun) on knowledge acquisition, situational interest, and academic emotions. Chapter 5 builds on this design by isolating a specific game feature—sound—and investigates its effects on the same outcome variables, while also incorporating behavioural engagement measures as outcomes. Chapter 6 presents the systematic review, which synthesizes research on digital games and intercultural outcomes. Finally, Chapter 7 integrates the findings across studies. It discusses insights from the development of the *Limes* learning environments, synthesizes the experimental results, relates them to findings from the systematic review, and outlines theoretical implications, methodological contributions, limitations and directions for future research.

.

## 1.1 Digital Games: Definitions and Conceptual Distinctions

Digital games (also referred to as video games; e.g., Mäyrä, 2008) are a widespread form of digital media, with an estimated number of 3.6 billion players worldwide (e.g., Newzoo, 2025). Digital games are interactive computer-based systems in which players engage with rule-based environments, pursue goals, and receive feedback on their actions. These systems typically involve structured challenges, player decision-making, and dynamic responses from the game environment that shape the progression of play. Within research on digital media and learning, digital games can be described as interactive systems in which players pursue goals under rules and receive feedback on their progress (Clark et al., 2016), often structured through cycles of challenge, response, and feedback (Plass et al., 2015).

Within this broad category, most digital games are developed primarily for entertainment purposes, that is, for player enjoyment, engagement, and leisure rather than explicitly defined instructional outcomes (E. A. Boyle et al., 2012). In contrast, serious games are games designed for purposes beyond entertainment (Deterding et al., 2011). Serious games aim to achieve specific outcomes such as education, training, or social awareness while maintaining core game characteristics such as goals, rules, and player interaction. Deterding et al. (2011) describe serious games as games that retain the structural properties of entertainment games but are developed with explicit functional purposes, including instructional or educational objectives. In educational contexts, these applications are often referred to as educational video games (Michael & Chen, 2006), emphasizing their use as instructional tools designed to support learning processes. Moreover, the term digital game-based learning (DGBL) (Prensky, 2001) is commonly used to describe the use of video games for educational purposes, encompassing a broad range of game types and learning applications (Carenys & Moya, 2016).

Building on these concepts, game-based learning refers to the pedagogical use of fully developed games within instructional settings. In game-based learning, digital games are integrated into learning environments in order to facilitate engagement with academic content through interactive challenges, feedback systems, and goal-directed tasks (Plass et al., 2015). In contrast, gamification involves integrating specific game design elements—such as points, badges, or leaderboards—into non-game learning environments to increase engagement and motivation (Deterding et al., 2011; Sailer & Homner, 2020).

In summary, digital games (or video games) constitute a broad technological medium

that is predominantly developed for entertainment, while serious games represent a subset designed for purposes beyond entertainment, including education, and are often referred to as educational video games in educational contexts. DGBL denotes the use of digital games for educational purposes regardless of their original design intent, whereas game-based learning refers to their instructional integration, and gamification involves the use of individual game design elements in non-game contexts. In this dissertation, the term DGBL is used as the primary umbrella term, while other terms are employed where they are conceptually more precise. The terms digital games and video games are used interchangeably throughout.

## 1.2 Digital Games as Learning Environments: Implications for Academic Outcomes

As digital technologies have become increasingly embedded in everyday life, they have also been increasingly adopted in educational contexts (e.g., OECD, 2015, 2021). In this context, digital games have been considered new forms of learning environment that can support educational processes in ways that differ from more traditional instructional formats. To understand the educational potential of DGBL, it is essential to examine how it influences different domains of academic outcomes. Research on DGBL commonly distinguishes between several interrelated domains, including cognitive outcomes, motivation, emotions, and behavioural engagement, which together capture important aspects of the complex ways in which learners respond to and potentially benefit from game-based environments (E. Boyle et al., 2019; Clark et al., 2016; Wouters et al., 2013). The remainder of this section discusses how digital games may influence these outcome domains, drawing on theory and reviewing empirical evidence.

*Cognitive outcomes* in DGBL research commonly refer to changes in learners' knowledge and cognitive skills that result from engagement with instructional materials (e.g., Wouters et al., 2013). In DGBL environments, cognitive processing is influenced by how information is presented and how learners interact with the system. Digital games can support cognitive outcomes by combining narrative structures, visual representations, interactive tasks, and feedback systems that actively engage learners with instructional content. This assumption is consistent with the ICAP (Interactive–Constructive–Active–Passive) framework (Chi & Wylie, 2014), which proposes that deeper forms of cognitive engagement, particularly constructive and interactive activities, lead to improved learning outcomes. Such interactive environments may facilitate deeper processing of information by encouraging exploration, problem-solving, and repeated engagement with learning materials (e.g., Chi & Wylie, 2014; Plass et al., 2015; Wouters et al., 2013).

Another relevant theory to explain the effect of DGBL on cognitive outcomes is the Cognitive Theory of Multimedia Learning (R. E. Mayer, 2014). According to the Cognitive Theory of Multimedia Learning (CTML), meaningful learning occurs when learners actively process information by selecting relevant verbal and pictorial information, organizing it into coherent verbal and pictorial representations, and integrating these with prior knowledge. These processes do not follow a fixed order but need to be coordinated for successful learning (R. E.

Mayer, 2014). Video games can support these processes by presenting information through complementary verbal and pictorial channels and by enabling interactive engagement with learning materials. At the same time, it has been cautioned that the presence of multiple simultaneous stimuli in digital environments can increase extraneous cognitive load if instructional design does not adequately guide attention. Instructional design principles such as signalling, structured progression, and scaffolding can therefore support learning by directing learners' attention toward relevant information and facilitating the integration of new knowledge into coherent mental models (Eitel et al., 2025; R. E. Mayer, 2014).

A meta-analysis by Clark et al. (2016), synthesizing results from 69 study samples and over 200 effect sizes, found that students in digital game conditions demonstrated significantly higher cognitive outcomes than those in non-game instructional conditions. However, the included studies were primarily conducted in domains such as mathematics, literacy, science, and psychology, while other areas, such as history, were largely underrepresented. Across studies measuring cognitive outcomes, the mean effect size was $g = 0.35$, indicating a small-to-moderate positive effect of video games on knowledge acquisition and related cognitive outcomes. These findings suggest that digital games can support learning when compared to more traditional instructional approaches such as lectures, reading tasks, or drill-and-practice activities. In another meta-analysis on simulation-based training studies, Sitzmann (2011) found that game-based approaches improved declarative and procedural knowledge compared to conventional instructional methods such as lectures, text-based materials, and computer-based instruction. Consistent with these meta-analytic findings, individual experimental studies have reported positive effects of digital games on cognitive outcomes. For example, Boeker et al. (2013) showed in a randomized controlled trial that game-based e-learning using the educational adventure game *Uro-Island* led to higher knowledge gains compared to conventional instructional approaches.

However, it has also been emphasized that cognitive benefits are not determined by the medium alone but depend strongly on the instructional design of the game, including features such as scaffolding, feedback, and structured learning tasks (e.g., Clark et al., 2016). An important strand of research in DGBL therefore examines such specific design elements rather than treating games as a uniform category. Here, particular attention has been given to features like feedback and instructional support that guides how learners engage with content and navigate the learning environment. For example, Mao et al. (2024), manipulated feedback type (dynamic vs. static) and task difficulty within a DGBL environment, showing that learning

outcomes varied depending on how feedback was structured and how it interacted with task conditions. By contrast, other design features remain comparatively underexplored (e.g., Plass et al., 2015; Wouters et al., 2013).

Overall, theory and research suggest that video games can support cognitive outcomes by promoting active, constructive, and interactive engagement with learning materials (e.g., Chi & Wylie, 2014; Plass et al., 2015; Wouters et al., 2013), while facilitating the selection, organization, and integration of information across multiple representations (R. E. Mayer, 2014). Building on this, two overarching complementary questions emerge: whether cognitive outcomes differ between game-based and non-game instructional formats, and how specific game design features influence cognitive outcomes within game-based environments.

*Motivation* represents a second important outcome domain in research on DGBL because it influences whether and how learners engage with instructional materials and sustain effort during learning activities (e.g., Eccles & Wigfield, 2002; Ryan & Deci, 2000). A widely used definition describes motivation as the process through which goal-directed activities are initiated and maintained (Schunk et al., 2014). In educational contexts, motivation therefore affects learners' willingness to engage with instructional tasks, persist when encountering difficulties, and invest cognitive effort in understanding new material (e.g., Linnenbrink-Garcia et al., 2016). Several theories have been developed to explain motivation in learning contexts. For example, Self-Determination Theory emphasizes basic psychological needs such as autonomy, competence, and relatedness as drivers of intrinsic motivation (Deci & Ryan, 2012). Other approaches focus on learners' value beliefs and success expectations, such as the situated expectancy-value theory (Eccles & Wigfield, 2020). Additional frameworks emphasize learners' interests (Hidi & Renninger, 2006), achievement goal orientations (Elliot, 1999), or self-concept and ability beliefs (Marsh & Craven, 2006). Although these theories differ in their specific foci, they converge in suggesting that motivation plays a central role in shaping learners' engagement with educational activities.

In the context of DGBL, motivation has received particular attention because video games often incorporate design elements—such as challenge, interactivity, feedback, and narrative structures—that are believed to support students' motivation (e.g., Plass et al., 2015; Wouters et al., 2013). Nonetheless, a meta-analysis by Wouters et al. (2013), synthesizing results from 39 studies on serious games, reported that game-based learning environments did not consistently produce higher motivation than conventional instructional conditions. Yet, several individual experimental studies have found positive effects of DGBL on learners'

motivation. For example, Hung et al. (2014) found that students who learned with a self-developed digital game-based instructional system reported higher levels of learning motivation than those receiving traditional instruction. Similarly, Papastergiou (2009) showed that students in a game-based learning condition reported higher motivation than those in a conventional computer-based instructional condition. These findings suggest that DGBL environments can stimulate motivational outcomes, but that further research is needed to identify the conditions and design features under which such effects emerge. Moreover, another potential explanation for the inconsistent findings on motivation reported by Wouters et al. (2013) is that different motivational constructs were combined. A more fine-grained approach that distinguishes between specific constructs, or between components within a construct, may therefore be more informative.

One construct that illustrates the importance of such distinctions is situational interest, which refers to a temporary state of attention and engagement triggered by features of the learning environment (Hidi & Renninger, 2006). According to the Four-Phase Model of Interest Development, situational interest may initially emerge as triggered situational interest, which is elicited by environmental features such as novelty, challenge, or interactivity. For learning to be sustained, however, this initial attraction must develop into maintained situational interest, where learners continue to engage with the content because they perceive it as meaningful or valuable (Hidi & Renninger, 2006; Linnenbrink-Garcia et al., 2010). This distinction is particularly relevant for DGBL, where engaging design elements may attract learners' initial attention but do not necessarily guarantee sustained motivation to engage with the learning content. While the distinction between triggered and maintained situational interest is theoretically well established (Hidi & Renninger, 2006; Renninger et al., 2019), research on DGBL usually does not seem to differentiate between these two types of interest.

In addition to cognitive and motivational outcomes, learning experiences are shaped by learners' *emotional responses* to instructional activities and learning tasks. The Control-Value Theory of Achievement Emotions proposes that emotions such as enjoyment and boredom arise from learners' perceptions of control over a task and the value they assign to it (Pekrun, 2006). According to the theory, these emotions influence learners' motivation, cognitive engagement, and the strategies they use during learning activities, thereby affecting learning outcomes (Pekrun, 2006; Pekrun & Linnenbrink-Garcia, 2014).

Video games frequently incorporate design elements that may influence these appraisals. Features such as autonomy in decision-making, interactivity, immersive narratives,

immediate feedback, and optimally challenging tasks may increase perceived control and task value, thereby fostering positive emotions such as enjoyment (Plass et al., 2015; Ryan et al., 2006). Conversely, when these features are absent or poorly implemented, control or value appraisals may be reduced, contributing to negative emotions such as boredom (Pekrun, 2006). Loderer et al. (2020; see also Plass et al., 2015) furthermore specifically discussed different "emotional design features" that may support enjoyment and counteract boredom, such as visual aesthetic design, musical scores, narratives, and incentive systems.

Empirical research on DGBL has increasingly begun to examine learners' emotional experiences. For example, Jackson and McNamara (2013) found that students learning with a game-based intelligent tutoring system reported higher levels of enjoyment as compared to students learning with a traditional tutoring system. Another study focused on emotional expressive behaviour as assessed using facial expressions and subjective feelings. The findings revealed increased positive as well as negative emotions in the game-based condition as compared to a non-game-based equivalent, which the authors attributed to the narrative elements, appealing visual aesthetics, and virtual incentives in the game-based version of the task (Ninaus et al., 2019). However, Wu et al. (2020) compared a DGBL condition with a static e-learning condition, using the latter as a non-game control group in which participants engaged with the same instructional content in a non-interactive format. They found that the DGBL condition did not have better affective experiences, measured as percentage of positive emotion, than the control condition.

Finally, learners actual *engagement* with learning tasks represents another key dimension for understanding learning in digital learning environments as it has been linked to performance (Fredricks et al., 2004; Lei et al., 2018; Martin & Borup, 2022), and constitutes an important outcome in its own right. Contemporary models of student engagement distinguish between behavioural, cognitive, and emotional dimensions of engagement, which together describe how learners participate in and experience instructional activities (Fredricks et al., 2004). Behavioural engagement refers to learners' (observable) participation in learning activities, including effort, persistence, time on task, and active involvement with instructional materials (Fredricks et al., 2004). Digital learning environments, including DGBL, provide particularly valuable opportunities to examine behavioural engagement because learners' interactions with instructional systems can be recorded automatically through system logs and behavioural trace data. Such data allow researchers to analyse how learners navigate instructional content, how much time they spend on specific tasks, and which learning materials

they access during learning activities (Winne, 2022; Wouters et al., 2013)[1]. Consequently, objective behavioural data can overcome problems commonly associated with self-report measures, such as social desirability, incomplete memories (e.g., Halem et al., 2020; Parry et al., 2021), by offering a more direct view of learners' interactions with digital learning environments.

Beyond self-reported behavioural engagement, research in DGBL has examined respective behavioural indicators and their relation to learning outcomes. From an engagement perspective, observable participation in learning activities such as effort, persistence, and active involvement with task elements are theoretically relevant because they reflect the extent to which learners engage with learning opportunities, which in turn can facilitate cognitive processing and learning (Fredricks et al., 2004; Von Keyserlingk et al., 2025). V. Shute and Ventura (2013) discussed that behavioural data collected during gameplay, such as how players solve problems and make decisions, can be used to assess their knowledge and skills without interrupting the learning process with separate tests. Empirical studies have shown that learners' in-game actions can provide meaningful information about their performance. For example, Serrano-Laguna et al. (2014) analysed in-game interaction data and player progression patterns to demonstrate how learning analytics can be used to monitor learners' activity and support assessment in digital games. More generally, research on digital learning environments shows that system-generated indicators of engagement (e.g., time spent studying relevant materials online) are associated with higher achievement (e.g., Louw et al., 2008; Schumacher et al., 2026; Spitzer, 2022).

At the same time, behavioural engagement can also be conceptualized as an outcome in research on DGBL where behavioural engagement indicators such as time-on-task, interaction frequency, or persistence are used to capture learners' involvement. In this perspective, behavioural engagement represents a meaningful educational outcome in its own right, reflecting learners' willingness to participate in and sustain interaction with the system. This is particularly relevant in digital and game-based environments, where continued participation is not guaranteed and must be actively supported by design (e.g., Deterding, 2012). From a game design perspective, tracking behavioural engagement and examining which elements (e.g., specific design features) influence it is important for evaluating whether the game and its features function as intended by supporting sustained player participation (i.e., encouraging

[1] Although this thesis adopts the term "behavioural engagement," some indicators may also be interpreted as self-regulated learning behaviours (Pintrich, 2004; Zimmerman, 2000).

players to keep playing). However, despite this interest in behavioural measures, relatively few experimental studies on DGBL have systematically examined effects on behavioural engagement.

### 1.1.1 Video Games for History Learning

History shapes how individuals and societies make sense of the present and can inform expectations about the future. Interpretations of the past provide frameworks through which contemporary social, political, and cultural issues are understood and evaluated, making historical knowledge a resource for public discourse (Rüsen, 2004). For individuals to engage with such interpretations, however, they must be able to understand past events within their historical contexts rather than interpreting them through contemporary assumptions. A central challenge in history education is enabling students to understand past events within the constraints of their historical contexts rather than interpreting them through contemporary assumptions. This problem is often described as presentism, the tendency to evaluate historical actions according to modern norms and values (Wineburg, 2001). Developing historical reasoning therefore requires learning environments that help students situate events within their social, political, and cultural contexts rather than merely memorizing isolated facts (van Drie & van Boxtel, 2008). Because historical reasoning depends on understanding the constraints and possibilities of a particular historical moment, instructional environments that simulate such contexts may offer advantages over purely descriptive formats (McCall, 2023).

DGBL has been proposed as one way to support this type of contextualized historical reasoning, as well as related and more readily measurable academic outcomes, such as knowledge about specific historical periods and learners' interest in history. By embedding information within rule-based systems and consequence-driven tasks, games allow learners to act within structured environments where decisions produce visible outcomes (Chapman, 2016; McCall, 2016). In this sense, historical games can create what McCall (2023) describes as "historical problem spaces", in which players act within rule-based environments representing past contexts. Through interaction with these systems, players may explore historical constraints and experiment with alternative actions and outcomes within the boundaries defined by the game's underlying historical models (Chapman, 2016; McCall, 2023).

However, these affordances do not guarantee educational effectiveness. Research on intrinsic integration shows that learning content must be embedded within core game mechanics to avoid superficial engagement (Habgood & Ainsworth, 2011). Many commercial titles prioritize entertainment mechanics that are only loosely connected to history-related learning objectives (Chapman, 2016; McCall, 2016). This tension highlights the need for purpose-built digital games that align historical content with core interactive systems under conditions that

permit empirical examination. If narrative or action elements are disconnected from historical content, learners may focus on surface features rather than developing contextual understanding.

Empirical research examining the effects of digital games on academic outcomes in the context of history learning remains relatively limited compared with other domains of DGBL (Clark et al., 2016). One frequently cited example is a mobile location-based history game investigated by Huizenga et al. (2009) in which secondary school students participated in a city-wide historical game set in Amsterdam. Students in the game condition navigated historical locations and completed tasks related to historical events. The results showed that students in the game condition achieved higher knowledge scores than those receiving traditional project-based instruction, while no significant differences were found with respect to motivation for the subject of history. These findings indicate that game-based historical environments can support knowledge acquisition, although the study does not isolate which specific elements of the game contributed to these effects. Other classroom-based studies have explored the integration of digital games into history instruction. For example, Yu et al. (2014) compared game-based instruction using the strategy game *Age of Empires* with traditional teacher-led instruction in a secondary school world history course. Students in the game-based condition engaged with historical content through gameplay, while the control group received conventional classroom instruction using lectures and standard teaching materials. The results indicated that students in the game-based condition reported higher learning outcomes than those receiving traditional instruction and also reported higher overall motivation toward the learning activity. Beyond such comparative studies, design-based research has also investigated how digital history games can be developed and implemented in classroom contexts. For instance, Çelik (2023) analysed the development and classroom use of a digital history game and highlighted the potential of interactive gameplay to support engagement with historical content.

Commercial historical games have also been examined in educational contexts. For example, Karsenti and Parent (2020) investigated the educational use of three titles from the *Assassin's Creed* series (*Assassin's Creed II*, *Assassin's Creed: Brotherhood*, and *Assassin's Creed: Revelations*) in secondary school history classes in Québec (Canada). The exploratory study combined classroom observations, teacher interviews, and student questionnaires to examine how the games were incorporated into teaching practices. Students reported high levels of interest and engagement, and teachers used game scenarios and historical representations to stimulate discussion and historical reflection. However, because the study focused primarily on

classroom use and reported experiences rather than controlled comparisons, it was not designed to and does not provide causal evidence regarding academic outcomes.

In addition, as compared to the studies providing experimental evidence in school contexts, research in higher education is even more limited. Some work has integrated commercial historical strategy games such as *Crusader Kings II* and *Europa Universalis IV* into university history courses, reporting that gameplay can support engagement with historical systems and contexts (e.g., Kuran et al., 2018). However, controlled experimental evidence on digital history games with university students seems widely lacking.

Overall, the empirical evidence base for digital history games remains fragmented. Many studies rely on exploratory classroom implementations, design-based research, or qualitative analyses rather than controlled experimental designs that allow causal conclusions about effects on academic outcomes. As a result, it remains difficult to determine the extent to which digital games systematically affect academic outcomes such as history knowledge, motivation, emotions, or behavioural engagement. This limitation highlights the need for carefully designed experimental studies on DGBL in the context of history learning that examine effects on academic outcomes.

### 1.1.2 The Case of Roman History

As the majority of work in this dissertation focuses on the development and experimental testing of a self-developed digital game on Roman history (*Limes*), the choice of historical context requires justification. The selection of Roman history, and in particular the Roman frontier (limes)[2] in the region of present-day Baden-Württemberg, was guided by both contextual and theoretical considerations.

From a contextual perspective, the initial target population of this dissertation project consisted of school students from Baden-Württemberg, a region historically shaped by the limes and its interactions with local populations. The use of a historically relevant period and location increases the meaningfulness of the learning material and helps learners connect new information to their prior knowledge (e.g., van Drie & van Boxtel, 2008; Wineburg, 2001). In addition, connecting historical content to learners' local and contemporary context has been argued to enhance the perceived relevance of history learning (Stefaniak et al., 2017). Roman history represents a central component of history education in many European curricula (Crivellari, 2010). In Germany, the limes is broadly covered in school curricula across federal states, reflecting its historical significance, for example, in curricula from Baden-Württemberg, Hessen, and Bavaria (Bayerisches Staatsministerium für Unterricht und Kultus, 2026; Hessisches Kultusministerium, 2002; Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2016).

From a theoretical and design perspective, the limes in this region is well documented through archaeological and historical research (Graafstal et al., 2021), providing a robust empirical basis for historically grounded representations of the past. This is particularly important for the design of video games, where accurate and meaningful integration of historical content are presumed to support effective learning (Lai & Hu, 2025).

Instead of testing *Limes* with school-aged students from Baden-Württemberg, however, it was later decided to carry out the experimental work with university students from Hessen, due to feasibility considerations and, given the scarcity of research on history video games in this target population (see Section 1.1.2), to examine the suitability of the game for this target group. While the specific historical setting of the game is located in present-day Baden-Württemberg, the limes extended across large parts of what is now southern and central

[2] The Roman limes refer to the frontier system of the Roman Empire, consisting of fortifications, watchtowers, and settlements; in present-day Germany, it formed part of the Empire's northern boundary.

Germany, including regions such as present-day Hessen and other areas along the Rhine River. This broader geographical scope reflects the historical significance of the limes as a frontier system that shaped multiple regions of the Roman provinces in Germania. As such, the broader historical context of the Roman Empire and its interactions with local populations remains regionally relevant, even if the direct local connection is less pronounced.

## 1.3 Beyond Academic Outcomes: Digital Games, Prejudice, and Intercultural Competence

While digital games are often examined in relation to academic learning outcomes, they have also been discussed as environments that can shape intercultural outcomes, such the development of intercultural competence and reductions in prejudice. This broader perspective aligns with global societal goals, as for example articulated in the United Nations Sustainable Development Goals (SDGs), including fostering global citizenship and intercultural understanding (SDG 4.7), reducing inequalities and discrimination (SDG 10), and promoting inclusive and peaceful societies (SDG 16) (United Nations, 2015).

Prejudice is commonly defined as a negative attitude or evaluation toward individuals based on their membership in a social group (e.g., ethnicity, cultural background) (Allport, 1954; Myers & Twenge, 2022; VandenBos & American Psychological Association, 2015), often involving affective, cognitive, and behavioural components such as negative feelings, stereotypes, and discriminatory tendencies. While prejudice captures evaluative responses toward members of other social groups, it does not account for individuals' ability to interact effectively in intercultural contexts. Accordingly, research has also examined intercultural competence as a distinct outcome domain. Intercultural competence is commonly defined as the ability to interact effectively and appropriately with individuals from different cultural backgrounds (Deardorff, 2006). Scholars often conceptualize this competence as multidimensional (e.g., Matsumoto & Hwang, 2013). For example, according to Byram (2020), the "intercultural speaker" is characterized by curiosity and openness toward other cultures, knowledge of social groups and cultural practices, the ability to interpret and relate cultural perspectives, skills for discovery and interaction, and critical cultural awareness. Deardorff (2006) further distinguishes between internal outcomes (e.g., empathy, flexibility) and external outcomes (e.g., appropriate and effective behaviour in intercultural interactions).

Several theoretical perspectives provide explanations for how prejudice and intercultural competence may change through mediated experiences. One prominent framework is intergroup contact theory, which proposes that positive interactions between members of different social groups can reduce prejudice and improve intergroup attitudes (Allport, 1954; Paolini et al., 2024; Pettigrew & Tropp, 2006). When contact occurs under favourable conditions—such as cooperation toward shared goals, equal status between participants, and institutional support—it can foster empathy, reduce prejudice and promote

more positive evaluations of outgroup members (Allport, 1954; Pettigrew & Tropp, 2006). Importantly, research suggests that similar processes may also occur in digitally mediated environments. A recent meta-analysis on digital intergroup contact indicates that interactions occurring through online communication, virtual environments, or digital media can reduce prejudice (Costa et al., 2024). Overall, these findings suggest that digital environments—including interactive media such as digital games—may provide opportunities for contact-like experiences with outgroup members or perspectives, thereby potentially reducing prejudices and fostering intercultural competence.

A second perspective stems from research on multiculturalism and diversity learning, which emphasizes the value of learning about different cultural perspectives and experiences (e.g., Rosenthal & Levy, 2010; Schwarzenthal et al., 2020). Although this line of research primarily examines regular educational contexts involving face-to-face interactions between students and teachers (e.g., Bardach et al., 2024) as well as work contexts (e.g., Holmes IV et al., 2021), similar principles may also apply to digital learning environments that expose learners to diverse cultural perspectives and narratives. The concept of a *cultural diversity climate* refers to the approaches through which schools and other social settings address cultural diversity, which are reflected in institutional practices, norms, and the interactions taking place in these settings (Bardach et al., 2024; Schachner et al., 2021). One important form of such a climate is a *multiculturalism climate*, which explicitly values cultural diversity and encourages students to learn about the experiences, practices, and perspectives of different cultural groups. Through such practices, educational environments may foster understanding and appreciation of cultural variation, and support the development of intercultural competencies and the reduction of prejudice (Bardach et al., 2024; Schwarzenthal et al., 2020). Digital games provide a potentially powerful context for engaging with these processes. For example, because games often involve role-taking, narrative interaction, and simulated social environments, they allow players to adopt and act within different roles and identities and engage in role-taking processes that foster perspective-taking (e.g., Bachen et al., 2012; Gee, 2007; Peng et al., 2010). These features may create conditions that resemble a multiculturalism climate, thereby providing a theoretical basis for why video games with such features may improve intercultural competence and reduce prejudice (see e.g., Schwarzenthal et al., 2020; for links between multiculturalism climate and intercultural competence).

At the same time, empirical research on the effects of digital games on intercultural outcomes remains limited and heterogeneous. Existing studies have investigated a range of

outcomes, including prejudice and components of intercultural competence, often without explicitly conceptualizing them as such. As a result, findings are mixed and may reflect variation in designs of the games and the context in which they are played as well as inconsistencies in how outcomes are conceptualized and operationalized (e.g., Bachen et al., 2012; Peng et al., 2010). This variability highlights the importance of examining both the design and implementation of games used and the specific mechanisms through which gameplay may influence intercultural competences and prejudices, as well as the need for greater conceptual clarity in how outcomes are defined and measured.

## 1.4 Substantive and Methodological Challenges in DGBL Research

Despite growing interest in DGBL, the field faces several substantive and methodological challenges. The following section outlines three key challenges that limit the interpretability of individual studies and the comparability of findings across studies.

### 1.4.1 Weak Comparison Logic in Game–Non-Game Studies

A recurring limitation in DGBL research concerns weak comparison logic between game-based and non-game instructional formats. Many studies compare digital games with non-game instructional conditions (e.g., lectures, reading materials, drill-and-practice, or hypertext environments), which often differ substantially in instructional design, level of interactivity, , pacing, feedback, and opportunities for engagement, complicating direct comparisons (e.g., All et al., 2014; Clark et al., 2016; Wouters et al., 2013). As a result, observed effects may reflect differences in structural affordances rather than properties inherent to the game format itself. Without alignment in learning opportunities and user interaction, it becomes difficult to determine whether performance differences and other academic outcomes are attributable to the medium, to variations in instructional design, or to differences in engagement intensity. This limits internal validity and complicates claims regarding the relative effectiveness of game-based formats.

### 1.4.2 Confounded Design Features in DGBL Research

A second challenge concerns the confounding of multiple design features within digital games. Digital games typically integrate narrative structure, audiovisual elements, interactivity, feedback systems, reward mechanics, and pacing within a single environment. In many experimental studies, several of these elements vary simultaneously across conditions, making it difficult to determine which specific feature is responsible for observed effects (e.g., R. E. Mayer & Moreno, 2003; Plass et al., 2015). Thus, when effects on academic outcomes are observed, it often remains unclear whether they stem from narrative design, audiovisual elements, feedback systems, or other design features. This complicates theory development, as conclusions about general “game effects” may conceal the impact of specific design elements, such as visual, auditory, or interactive supports. The lack of feature-level isolation has been repeatedly identified as an obstacle to cumulative research and to the identification of underlying learning mechanisms in DGBL (e.g., Plass et al., 2015).

### 1.4.3 Fragmentation of Outcome Constructs and Research Agendas

Beyond these general design and comparison issues additional challenges arise in specific application domains. Specifically, research on digital games and intercultural outcomes is characterized by fragmentation in both outcome constructs and research agendas. In particular, key constructs—such as intercultural competence—are defined and operationalized inconsistently across studies. This conceptual heterogeneity limits the interpretability of individual findings and reduces the comparability of results across studies.

Furthermore, research agendas in this area are fragmented, as existing studies adopt heterogeneous approaches, focusing on different aspects of digital games. For example, some investigate comparisons between instructional formats (e.g., video game versus text; H.-P. Chen et al., 2010), others analyse particular gameplay features such as avatar embodiment or character interaction (e.g., Behm-Morawitz, Hoffswell, et al., 2016), while additional work focuses on differences between various types of video games (e.g., Saleem & Anderson, 2013). Because these lines of research target different mechanisms and are rarely systematically integrated, they provide limited cumulative insight and hinder the development of a coherent understanding of which aspects of video games influence intercultural outcomes.

## 1.5 Positioning and Scope of the Present Dissertation

The preceding sections identified three central challenges in research on DGBL: weak comparison logic between game and non-game formats, insufficient isolation of specific design features, and fragmentation in research agendas and outcome constructs. To address these challenges, the present dissertation combines design development, experimental investigation, and systematic synthesis within a coherent research program. Central to this approach are two purposefully designed learning environments: the educational video game *Limes* and a structurally aligned interactive text application. These environments enable controlled comparisons between formats while holding key elements constant, providing the basis for an experimental study on format effects. In addition, the use of a self-developed game allows for the systematic manipulation of specific design features; this is employed in a second experimental study to examine the effects of audio elements as a selected design feature. Both experimental studies focus on cognitive, motivational, emotional, and behavioural engagement outcomes in the context of digital history learning. Complementing these studies, the dissertation includes a systematic review on digital game effects on intercultural outcomes (prejudice and intercultural competences), which deliberately extends beyond the domain of history learning to address fragmentation in outcome constructs and research agendas at the level of the broader DGBL literature by synthesizing a heterogeneous body of work across different research lines.

The present dissertation draws on diverse perspectives, including education science, history learning, and social and cultural psychology and proceeds from the premise that progress in DGBL research requires methodological precision and clear alignment between theoretical assumptions and empirical design. Also, rather than treating games as a single format, the dissertation conceptualizes them as structured learning environments composed of specific design features whose effects must be examined under controlled conditions. Overall, this integrated approach allows for (a) more interpretable comparisons between learning formats, (b) feature-level analysis of game design elements, and (c) a systematic synthesis of heterogeneous findings on intercultural outcomes, thereby reducing fragmentation in outcome constructs and research agendas. The following chapter specifies the overarching aims and research questions guiding this program.

# 2

# AIMS AND RESEARCH QUESTIONS

# 2 Aims and Research Questions

Having established the theoretical foundations and methodological and substantive limitations of existing DGBL research, the present chapter specifies the aims and research questions guiding the dissertation. This dissertation is guided by the following overarching research questions (RQ):

RQ1: How can digital games for history learning and their corresponding control conditions be designed to enable interpretable comparisons between instructional formats and isolate specific design features?

RQ2: What roles do digital games and their selected design features play for cognitive, motivational, emotional, and behavioural engagement outcomes in history learning?

RQ3: How is research on digital games and intercultural outcomes structured in terms of outcome constructs and research agendas, and what patterns can be identified across this heterogeneous body of work?

## 2.1 Strengthening Comparison Logic: Format and Framing

The first component of the research program addresses weak comparison logic in game–non-game research (All et al., 2014; Clark et al., 2016; Wouters et al., 2013). Many previous studies contrasted interactive games with structurally impoverished control conditions, thereby limiting internal validity. To overcome this limitation, a structurally aligned interactive text-based application was developed alongside the educational history game *Limes*, documented in two design papers: Paper 1 (*Limes* video game): Viccari et al., 2024; Paper 2 (*Limes* text application): Viccari & Bardach, 2025). Both formats share narrative structure, informational content, and progression systems, enabling a controlled comparison of delivery format while holding learning opportunities constant.

Building on these developments, Paper 3 presents an experimental investigation of format effects by comparing the educational video game *Limes* with the parallel text application. In addition, this study experimentally manipulates the instructional framing (e.g., whether the goal is to learn or have fun, Erhel & Jamet, 2013). Instructional framing may represent an additional source of variation in game–non-game comparisons and may interact with instructional format. Paper 3 thus implements a 2 × 2 experimental design (Game vs. Text × Learning vs. Fun instructional framing). Within this design, the study examines whether format, framing, and their interaction affect knowledge acquisition, situational interest (triggered and maintained), and academic emotions (enjoyment, boredom) in digital history learning.

## 2.2 Isolating Design Features: Sound as a Targeted Manipulation

A second line of inquiry responds to the confounding of multiple design features in digital games (e.g., R. E. Mayer & Moreno, 2003; Plass et al., 2015). Because digital games typically integrate narrative, interactivity, feedback, and audiovisual elements simultaneously, many experimental studies cannot determine which specific affordances drive observed effects. Paper 4 builds directly on the same game environment (*Limes* video game) and outcome measures as Paper 3 but isolates a single design feature: sound. Specifically, the audio condition includes narration (character voices and voice-over explanations of additional historical content) and ambient background sound. Two versions of *Limes* are compared that differ only in auditory presentation. Grounded in principles of multimedia learning and game design (R. E. Mayer & Moreno, 2003; Plass et al., 2015), this manipulation enables a focused examination

of how auditory elements affect knowledge test performance, situational interest, and academic emotions under controlled conditions. Moreover, Paper 4 includes behavioural engagement (in-game behaviours) as an additional outcome.

## 2.3 Clarifying Outcome Constructs and Organizing Research Agendas

The third component of the dissertation addresses fragmentation at the level of outcome constructs and research agendas. Paper 5 (systematic literature review) synthesizes research on digital game effects on intercultural outcomes, where constructs, particularly intercultural competence, have often been inconsistently conceptualized. In addition, Paper 5 addresses fragmentation in research agendas by organizing heterogeneous lines of work according to differences in game design, usage context, and underlying mechanisms.

## 2.4 Integrative Contribution

Taken together, by (a) strengthening comparison logic, (b) isolating individual design mechanisms, and (c) clarifying construct and research agenda fragmentation, the dissertation refines the evidential basis of DGBL research, with a particular focus on history learning. Rather than advancing undifferentiated claims about "game effects," the work specifies the conditions under which particular formats and design features influence critical outcome variables. In doing so, it contributes to greater methodological precision and conceptual clarity in DGBL research.

.

# 3

# METHODOLOGICAL FOUNDATION: DESIGN PAPERS

# 3.1

# Paper 1

Viccari, A., Göllner, R., Hahn, J.-U., & Bardach, L. (2024). Limes (the roman frontier): Developing a video game for history learning. *Proceedings of the European Conference on Games Based Learning, 18(1), 839–847*. https://doi.org/10.34190/ecgbl.18.1.2833.

**Abstract**

The use of gaming technologies in education holds great promise to provide engaging learning experiences. Whereas video games have been successfully used to foster learning in several school subjects, the current understanding of the potential of video games for history learning is still limited. This paper details the development of *Limes*, a 2D top-down educational game created in Unity and rendered in pixel art. The game targets 5th/6th grade secondary school students and is set in the third century AD in southern Germany, a period characterized by political and social instabilities. *Limes* digs into the complex relations between the Roman Empire and the Alemanni tribe, one of the most powerful Germanic tribes of this period. The player can experience this pixelated world through the eyes of an Alemanni child, who forms a friendship with a Roman centurion. The unfolding narrative places the young protagonist in the precarious balance of peace and conflict, and the player needs to navigate the diplomatic tensions of the era. The gameplay consists of five levels, including primary missions and optional secondary missions which unlock special features (e.g., puzzles, quizzes). *Limes* incorporates a variety of game mechanics, ranging from running and interacting with characters to stealth missions, throwing rocks, and horse racing. These dynamics are integrated into the historical narrative of the game, enriching the player's experience. The game development process combined insights and methods from multiple disciplines, including game-based learning, educational, social, and cultural psychology, education sciences and pedagogy, and history education. Moreover, the game capitalizes on Unity's wide array of development tools, which provide extensive customization possibilities, and employs pixel art, facilitating an engaging environment while maintaining a cost-conscious development process. Initial tests and expert feedback highlight the promise of *Limes* as a game that leverages historical accuracy and innovative educational tools to foster a deeper understanding of Roman history. Next steps include empirical studies with students to test the game's effectiveness (e.g., in improving interest in and knowledge of Roman history).

*Keywords:* Educational gaming, history education, game-based learning, intercultural understanding, Unity development

## 1. Background

The application of digital technologies in education and cultural heritage has proven to be a promising avenue for engaging diverse audiences through immersive, dynamic, and personalized experiences (Arrigoni & Galani, 2019; Kelpšienė et al., 2023). In particular, video games have emerged as innovative tools in these fields, as pilot projects within museum contexts have shown their potential to create novel engagement pathways and make historical knowledge accessible to broader audiences, including those without prior background knowledge (e.g., Ropeik, 2014; Johnson et al., 2015; Viccari, 2016; Hiriart, 2019; Harrington, 2023). History education is an auspicious context for video game interventions, engaging learners in reflecting on the significance of the past and its impact on the present. For a long time, history has been regarded as a subject that only required memorizing facts and data from the past (Nachtigall et al., 2022; Van Drie & Van Boxtel, 2008). However, in the last decades, it has been argued that history is not about accepting or rejecting facts uncritically but is viewed as an essential aspect of participation in a democratic society (Nachtigall et al., 2022; Van Drie & Van Boxtel, 2008). For instance, Barton and Levstik (2004) argue that history should promote reasoned judgment about important human matters. The importance of history goes beyond the learning of historical facts and encompasses several processes that imply reflecting, judging and deliberating over a historical event (Nachtigall et al., 2022; Van Drie & Van Boxtel, 2008). A number of terms have been adopted for describing these ideas, such as historical literacy (e.g., Lee, 2005, 2007; Perfetti et al., 1995), historical thinking (e.g., (Schreiber et al., 2006; VanSledright & Frankes, 2000; Wineburg, 2001) historical consciousness (Van Drie & Van Boxtel, 2008) and historical reasoning (e.g., Kuhn et al., 1994; Leinhardt et al., 1994). Video games, with their capacity to create narrative learning environments, are particularly well-suited to enhancing history education (Akkerman et al., 2009). People tend to interpret historical events in different ways, based on social-economic or personal ideas (Carretero et al., 1997). Especially students often use modern contexts and view the past through the lens of the present, which is an important source of misunderstandings of history (Akkerman et al., 2009). Situating a historical phenomenon in a historical context, which can be achieved by narratives, is a key activity in history education and can help overcome such misunderstanding. Narratives organize historical information in a coherent and contextualized way (Akkerman et al., 2009; Falk & Dierking, 2000; Liu, 2022; Sugiyama, 2001), and narratives can easily be integrated into video games (Akkerman et al., 2009). . The use of history as a main theme for educational video games has been described as particularly promising by McCall (2016), who provided a guide for new and experienced educators who want to engage their students towards history by

using existing historical commercial video games, such as Assassin Creed, Civilization and Total War. Nonetheless, the current understanding of the potential of video games for school students' history learning is still limited. Moreover, history video games that are designed to closely align with specific learning objectives outlined in the curriculum and that could thus easily be integrated into regular classroom teaching by history teachers are scarce.

In response to this gap, the present paper details the development of *Limes*, a 2D top-down history game created in Unity and rendered in pixel art, specifically designed for use in 5th/6th grade history classes. The following sections illustrate the steps taken to bring *Limes* to life and how the game integrates educational theories to foster a deeper understanding of Roman history among young learners.

## 2. Outline of the Project: How *Limes* has Come to Life

### 2.1 Concept

The incorporation of historical themes in educational video games has been identified as highly beneficial (e.g., McCall, 2016). Importantly, a balanced integration of narrative and gameplay is particularly crucial in educational video games where engagement and learning must coincide (Booker, 2006; Schell, 2014). In addition to consulting key video game design literature (e.g., Booker, 2006; Schell, 2014) an important first step of the development of *Limes* involved understanding what the main target group (German school students) needed to learn about Roman history. As the Roman Empire is an integral part of the 5th/6th grade curriculum in German secondary schools, it was decided to focus on these age groups and grade levels. By understanding the main topics covered in the curriculum, it was possible to pick a historical period that was well-suited for a video game. As the name suggests, one of the main topics in *Limes* is the limes itself, referring to the functions of the ancient Roman border in Germany and across the empire. Other important themes were the relationships between the Roman Empire and the Germanic folks and a focus on Germanic tribes' history, with particular attention to the Alemanni. The educational objectives of *Limes* are to foster students' motivation (e.g., interest in history), understanding of Roman-Germanic relations and knowledge about Roman history, historical empathy, critical thinking, and intercultural outcomes. These objectives guided the development of the narrative and gameplay, ensuring that each element contributes to meaningful learning outcomes. The narrative of *Limes* is set in third-century AD southwestern Germany and follows an Alemanni child, caught in a plot that could ignite a war. This storyline provides a lens for exploring historical themes, through the child's journey and interactions with a Roman centurion named Ariovist. This storyline is the result of balancing the aspects of game

design, narrative and history education.

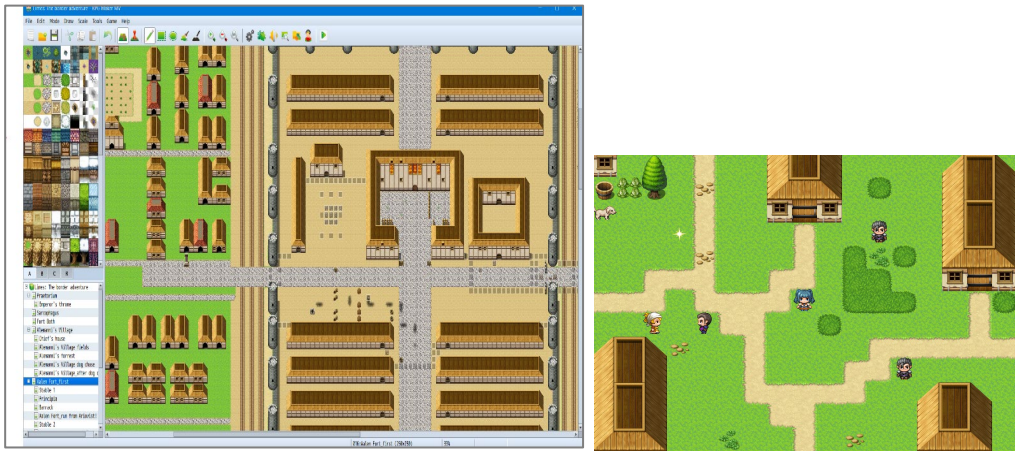
The next step was to figure out the best and most effective way to bring this historical period and its characters to life. The initial idea was to develop a 3D video game with a cartoony style; this choice was connected to two main reasons. Firstly, it is easier and less time-consuming to develop cartoon-like characters than real-person characters and 3D animations have been found to engage a wide audience (Bedrina, 2018; Matsangou, 2018). However, after developing a 3D character (see Figure 1), it was noticed that 3D development was not the right choice for the development of *Limes*, which is developed as part of the PhD project of the first author. Given the limited time frame of a PhD, the game was moved to 2D, allowing the creation of faster assets, such as building and natural features and game characters. The decision to switch to 2D was also corroborated by pretests conducted at the Technical University in Stuttgart (Hochschule der Medien); these tests were conducted with game design students and professors to obtain feedback about the game's story. To allow them to play something, the first author developed a very simple 2D video game using a software called RPG-Maker (see Figure 2). Some feedback suggested that a 3D historical video game may raise unrealistic expectations due to comparisons made with some of the current commercial historical video games, such as the well-known Assassin´s Creed series. In conclusion, the decision to make a 2D video game was influenced by project resources and timeframe, and also aimed to avoid creating a product that raises too high expectations from the outset. In the next sections, the game development and its features will be explained in more detail.

**Figure 1**

*3D model representing a Roman Centurion developed in ZBrush.*

**Figure 2**

*Screenshot of the software used to make a pretest of the video game (left side) and gameplay screenshot from RPG-Maker (right side).*



## 2.2 Levels and Codex System

*Limes* consists of five levels that guide players through the narrative journey of the Alemanni child protagonist during Roman times. Each level—The Village, The Fort, The Limes, The Unknown, and The Return (see Figure 3)—incorporates distinct historical themes that provide insights into the Roman and Alemanni history (Fuchs, 1998; Garnsey & Saller, 2014). After completing each level, players receive a performance-based score ranging from 1 to 5 stars. This rating reflects how many additional tasks the player has completed within the level.

**Figure 3**

*The five game levels of Limes, from left to right: Level 1 (The Village), Level 2 (The Fort), Level 3: The Limes, Level 4 (The Unknown), Level 5 (The Return).*

Level 1: *The Village* introduces players to the story's protagonist, giving some initial insights about the Alemanni people and a brief introduction to the Roman Empire as a political entity. This sets the stage for the cultural contrasts that follow in the next levels. This level introduces other important characters, such as the child's father, a powerful and wise Alemanni chef, and his advisor, Lüger, who will play a more important role towards the end of the story. The real adventure starts at the end of the level when the child gets kidnapped by Roman soldiers.

Level 2: *The Fort* details the architectural and organizational structure of a Roman military outpost, highlighting the strategic importance of such locations, especially close to the empire's borders. This level introduces one of the most important characters in the video game, Ariovist. He is a stoic and proud Roman centurion tasked with bringing the child back to his village. Although this task may seem like an act of kindness, it is initially merely done to avoid an open conflict with the powerful Germanic tribe. His sentiments will be clear from the beginning, creating a confrontational relationship with the child and highlighting how much Romans and Germans despise each other. This relationship will become clearer as the game progresses, giving important insights into cultural differences and their complexities.

Level 3: *The Limes* explores the Roman Empire's border defences, giving players insights into its expansion and interaction with neighbouring tribes. During their journey, the duo unexpectedly falls into two traps set by adversaries seeking to capture the child for a ransom of gold and silver. In a turn of events, the child bravely rescues Ariovist from a dangerous situation. Similarly, Ariovist demonstrates his loyalty by risking his life to protect the child. These acts of courage and sacrifice significantly deepen their relationship, highlighting themes of friendship and trust despite their cultural differences and memberships to different groups.

Level 4: *The Unknown* takes place in a remote Germanic village, reached after the

protagonist and Ariovist unintentionally stray from the limes into a dense forest. Lost and isolated, they stumble upon this village, a setting that challenges Ariovist's preconceived notions and exposes him to the culture of the Germans. This level serves not only as a geographical detour but as an important point for character development and narrative depth. During their time in the village, they encounter Lüger, who happens to be the architect behind the earlier kidnapping attempts of the child. After capturing the child and the centurion, Lüger reveals that his deepest ambition is to ignite a war against the Romans, driven by his belief that such a conflict would elevate the glory of the Alemanni and cement his legacy as a formidable leader. These revelations will bring the child and Ariovist even closer, laying a solid foundation for a friendship that transcends their initial differences. At the end of the level, the duo manages to escape from Lüger's control and sets off to warn the child's father about his treacherous advisor.

Level 5: *The Return* marks the dramatic conclusion of the story. Ariovist makes a noble sacrifice to protect the child from capture, and the child is reunited with the father. Despite Ariovist's selfless gesture, the father remains unwilling to reciprocate by intervening to rescue Ariovist. However, the child decides to confront the danger alone, risking everything to rescue his Roman friend. In a climactic showdown, the child confronts and ultimately overcomes Lüger, putting an end to his malevolent plans. This courageous act not only saves Ariovist but also cements their friendship, which has been forged and strengthened through challenges and mutual sacrifice. The level concludes with a reflection on how the protagonist and Ariovist have grown and how their perspectives on each other and the Roman and Germanic cultures, respectively, have evolved, shaped by the experiences they have shared during their journey.

Each level has a mandatory main mission and a non-mandatory secondary mission. The secondary missions are meant to give players a chance to experience extra missions with different game dynamics and gather more historical information. Secondary missions provide players with opportunities to explore diverse game dynamics, such as stealth operations, battling enemies and solving puzzles. Additionally, players can experience unique activities not available in primary missions, such as participating in horse races. As the player's performance is assessed based on the number of extra tasks completed during a level, completing secondary missions rewards players with higher scores at the end of each level.

A central feature of *Limes* educational architecture is the codex system, an innovative feature that allows players to unlock historical information through interactions with game characters. This system enriches the narrative and serves as a dynamic educational tool,

enabling players to access detailed explanations of historical figures and events at any point in the game. This information is accompanied by images and customized historical maps designed specifically for this game (see Figure 4). The engagement with the codex can be tracked to analyse its impact on learning outcomes (e.g., testing whether players who spend more time reading codex entries learn more).

**Figure 4**

*Gameplay screenshot of the codex system of Limes (left side) and historical map developed for the Limes codex system using a software called QGIS (right side).*


## 2.2 Game Development in Unity

The development of *Limes* was facilitated primarily through Unity, a powerful cross-platform game engine renowned for its flexibility and wide usage in indie and professional game development. Although RPG Maker might seem like a simpler option Unity was chosen for its versatility and the greater control it offers over the game's development. In addition, Unity possesses a huge asset store, which can facilitate the use of numerous types of assets, accelerating the creation of low-budget video games, such as *Limes*. The work in Unity can be divided into three sections: asset creation, coding in C#, in-build Unity development and voice integration with ElevenLab.

1. Asset creation employed pixel art graphics, which was achieved by using a software called Aseprite. This pixel art tool is specifically designed to create animations and sprites, allowing for the portrayal of historical themes with artistic clarity (see Figure 5). Pixel art not only invokes the nostalgia of classic games but also serves as a visually simplistic yet effective medium to convey complex historical narratives, ensuring that the educational content is accessible and engaging to learners (Egenfeldt-Nielsen et al., 2016; Messaris & Humphreys, 2007; Squire et al., 2011).

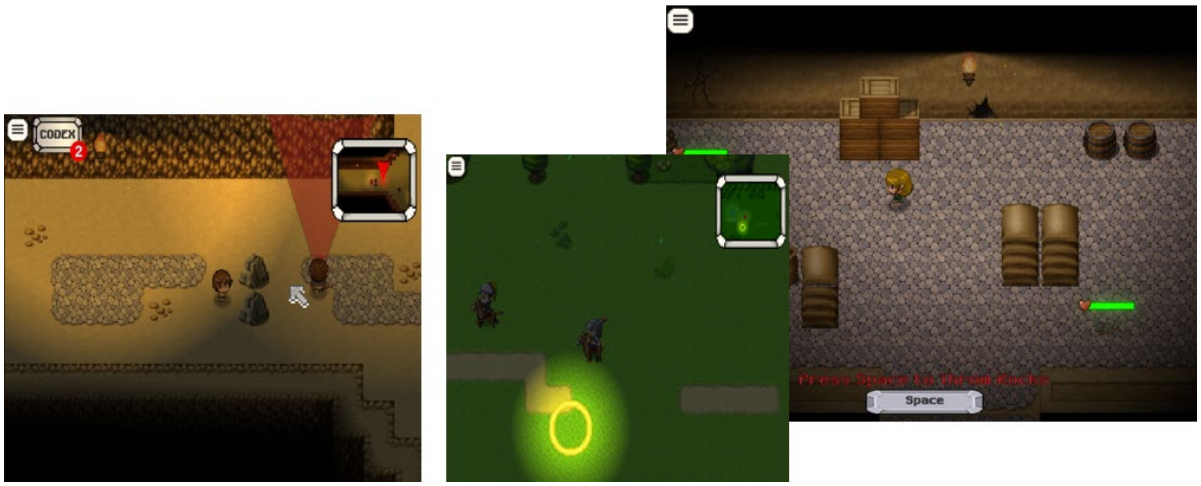
**Figure 5**

*Examples of characters used in Limes.*

2. Central to the development of the video game was the use of C#, Unity's primary programming language. The flexibility and object-oriented capabilities of C# enabled the creation of complex gameplay mechanics and interactive learning modules tailored to the historical content. Scripts written in C# facilitated the integration of a complex dialogue system with decision-making scenarios, which are vital for engaging students in active learning and critical thinking about historical events (Lin et al., 2023). Scripting was essential in developing the different game dynamics included in the video game. The game dynamics of *Limes* are moving in the environment and interacting with game characters through dialogues. During a mission, the player will be allowed to throw rocks at targets, engage in horse races and be involved in stealth missions, where it will be important not to be seen by the enemy (see Figure 6). Moreover, C# was vital in creating essential game features such as the codex system and the game database.

**Figure 6**

*Gameplay screenshot during a horserace mission (middle), gameplay screenshot during a throwing rock mission (right side), and gameplay screenshot during a stealth mission (left side).*




3. The build process within Unity involved layering the pixel art assets with the coded functionalities to assemble the game's levels and interfaces. This phase optimised the interaction between graphical elements and script-driven behaviours, ensuring seamless gameplay and educational delivery. Unity's Scene Management and Prefab systems were extensively utilized to manage and reuse game objects across different levels efficiently, enhancing development time and runtime performance. Another important feature used during this phase was the Unity asset store, which contains many free and paid assets. The asset store provided important features used within the game, such as terrain tiles, dialogue system, environment noises and music. This helped create the five levels of *Limes* effectively, saving development time.

4. Integrating voice narration in *Limes* involved employing ElevenLabs, an advanced AI tool that converts text to audio. Incorporating voice acting was not merely an aesthetic choice but a strategic enhancement to the game's immersive quality. By giving the characters distinct voices, the game becomes more engaging and accessible, helping to bring the historical narratives and personalities to life. This auditory dimension significantly enriches the player's experience, making the educational content more enjoyable and memorable (Dewatri et al., 2023).

*Limes* uses MongoDB, a flexible and scalable NoSQL database, to manage and store gameplay data effectively. This integration with Unity allows for seamless real-time data capture, which is crucial for the dynamic environment of the game. The database records a variety of metrics, such as levels completed, individual-level performance, total playtime, and specific interactions with hidden game features like quizzes and puzzles that are only accessible to players who perform very well in the game (a feature often used in commercial video games).

Furthermore, it tracks interactions with the codex, including how many times and how long codex entries are accessed, providing insights into the educational engagement of players.

### 2.3 A Multi-disciplinary Approach to Game Development

The game development process combined insights and methods from multiple disciplines, including game-based learning, psychology, history education, and education sciences and pedagogy. The educational objectives of *Limes*—fostering students' motivation (e.g., interest in history and the Roman Empire), knowledge about Roman history, historical empathy, critical thinking, and intercultural outcomes (e.g., positive diversity beliefs)—guided the integration of these disciplines into the game's design. A strong adherence to game-based learning principles ensured a balance between the need to cover subject matter and engaging gameplay (Plass et al., 2015). In addition, insights from educational psychology, for example, on how to trigger and maintain students' situational interest as precursors of developing an enduring personal interest in history and the Roman Empire (Hidi & Renninger, 2006), influenced the design of tasks in the game. Also, based on cultural and social psychology research, the game emphasizes the value of cultural diversity and learning about different cultures (here: Romans and Alemanni) (Bardach et al., 2024). Furthermore, principles of the well-established social psychological intergroup contact theory (Pettigrew & Tropp, 2006) were integrated. Specifically, intergroup contact theory states that positive contact between members from different groups can reduce prejudice and foster tolerance (Pettigrew & Tropp, 2006). The friendship between the child protagonist and Ariovist is an excellent example of the potential for positive intergroup contact. From education science, generic principles of instructional quality (e.g., the importance of cognitively stimulating learning content) were adopted. Based on insights from history education and didactics, the narrative emphasizes historical empathy and perspective-taking, and helps players understand complex historical events. It combines engaging storytelling with careful history examination, improving player's history knowledge and critical-thinking skills (Wessell, 2017).

### 2.4 Initial Tests and Expert Feedback

Following the development of the initial version of *Limes*, the game was tested by a diverse group of participants, including academics with backgrounds in education, psychology, and history, as well as avid video gamers. This broad spectrum of testers was chosen to gather insights and ensure the game's educational and entertainment values were optimized across different user groups. Overall, players showed great interest in the game and evaluated it very positively. The feedback received was instrumental in refining the game's design and

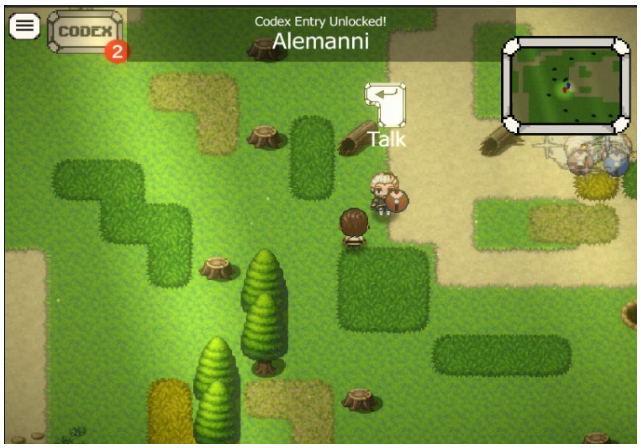
functionality. For example, participants highlighted the need for clearer instructions regarding the use of the codex system. In response, the game was updated to include more intuitive guidance on how to access the codex, accompanied by a notification system (see Figure 7) that alerts players when new information is available. This adjustment aimed to enhance the educational experience by ensuring players are fully aware of the learning resources at their disposal.

Another significant area of enhancement was the gameplay mechanics and dialogue accuracy. Testers identified several glitches that affected the game's flow and immersion. These issues were addressed, improving the overall smoothness and playability. Additionally, considering the game's bilingual nature—available in both English and German—special attention was given to the large volume of dialogues and their corresponding audio recordings. Translation and audio sync errors were corrected to ensure consistency, giving all players the same quality of content and storytelling. Overall, the insights gained from this phase have not only improved the game's educational efficacy but have also enhanced its engagement and usability, making it a more powerful tool for learning and enjoyment.

An important next step will involve in-depth pilot tests with secondary school students as the target group of *Limes*. After that, experiments in the lab will be conducted to test the usability of the game and effects of specific game elements. Then, it is planned to conduct a large-scale evaluation study in history school classes to gain insights into the effectiveness of the game in improving central outcomes (e.g., interest in history and the Roman empire, knowledge on the Roman Empire and history understanding, positive diversity beliefs).

**Figure 7**

*Gameplay screenshot showing the implementation of a notification system to unlock codex entries.*




## 4. Conclusions

To conclude, *Limes* leverages historical accuracy and innovative educational tools. *Limes* provides an engaging gameplay experience and a rich, interactive learning environment, effectively combining insights from multiple disciplines to achieve its educational objectives. Initial tests and expert feedback highlight the game's potential as a valuable resource in history education. Empirical studies with school students are now crucial to evaluate the game's educational effectiveness and inform scaling strategies for its broader use in (German) secondary school history classes.

**References**

Akkerman, S., Admiraal, W., & Huizenga, J. (2009). Storification in History education: A mobile game in and about medieval Amsterdam. *Computers & Education*, *52*, 449–459. https://doi.org/10.1016/j.compedu.2008.09.014

Arrigoni, G., & Galani, A. (2019). Digitally enhanced polyvocality and reflective spaces. *European Heritage, Dialogue and Digital Practices*, 37–61. https://doi.org/10.4324/9780429053511-3

Bardach, L., Röhl, S., Oczlon, S., Schumacher, A., Lüftenegger, M., Lavelle-Hill, R., Schwarzenthal, M., & Zitzmann, S. (2024). *Cultural Diversity Climate in School: A Meta-Analytic Review of Its Relationships With Intergroup, Academic, and Socioemotional Outcomes*. OSF. https://doi.org/10.31219/osf.io/fvjks

Barton, K. C., & Levstik, L. S. (2004). *Teaching history for the common good*. Lawrence Erlbaum Associates Publishers.

Bedrina, K. (2018, May 30). *Why Animated Video is the Best Strategy for Your Social Media*. Dreamgrow. https://www.dreamgrow.com/animated-video/

Booker, C. (2006). *The Seven Basic Plots: Why We Tell Stories*. Bloomsbury UK.

Carretero, M., López-Manjón, A., & Jacott, L. (1997). Explaining historical events. *International Journal of Educational Research*, *27*, 245–253. https://doi.org/10.1016/S0883-0355(97)89732-7

Dewatri, R. A. F., Aqthar, A. Z. A., Pradana, H., Anugerah, B., & Nurcahyo, W. H. (2023). Potential Tools to Support Learning: OpenAI and Elevenlabs Integration. *Southeast Asian Journal on Open and Distance Learning*, *1*(02), Article 02. https://odelia-journal.seamolec.org/index.php/current/article/view/21

Egenfeldt-Nielsen, S., Smith, J. H., & Tosca, S. P. (2016). *Understanding Video Games: The Essential Introduction* (3rd ed.). Routledge.

Falk, J. H., & Dierking, L. D. (2000). *Learning from Museums: Visitor Experiences and the Making of Meaning*. Rowman & Littlefield.

Fuchs, K. (1998). *Die Alamannen [The Alemanni]*. Theiss in Herder.

Garnsey, P., & Saller, R. (with Elsner, J., Goodman, M., Gordon, R., & Woolf, G.). (2014). *The Roman Empire: Economy, Society and Culture* (2nd ed.). University of California Press.

Harrington, C. (2023). Experience in Videogame Display: An Extension of the Matrix Model. *Games and Culture*, 15554120231209801. https://doi.org/10.1177/15554120231209801

Hidi, S., & Renninger, K. A. (2006). The Four-Phase Model of Interest Development. *Educational Psychologist*, *41*(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

Hiriart, J. (2019). *Gaming the past: Designing and using digital games as historical learning contexts*.

Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A. (2015). The NMC Horizon Report: 2015 Museum Edition. In *New Media Consortium*. New Media Consortium. https://eric.ed.gov/?id=ED559371

Kelpšienė, I., Armakauskaitė, D., Denisenko, V., Kirtiklis, K., Laužikas, R., Stonytė, R., Murinienė, L., & Dallas, C. (2023). Difficult heritage on social network sites: An integrative review. *New Media & Society*, *25*(11), 3137–3164. https://doi.org/10.1177/14614448221122186

Kuhn, D., Weinstock, M., & Flaton, R. (1994). Historical reasoning as theory-evidence coordination. In M. Carretero & J. F. Voss (Eds.), *Cognitive and instructional processes in history and the social sciences* (pp. 377–401). Lawrence Erlbaum Associates Publishers.

Lee, P. (2005). Historical Literacy: Theory and Research. *History Education Research Journal*, *5*. https://doi.org/10.18546/HERJ.05.1.05

Lee, P. (2007). From National Canon to Historical Literacy. In M. Grever & S. Stuurman (Eds.), *Beyond the Canon: History for the Twenty-first Century* (pp. 48–62). Palgrave Macmillan UK. https://doi.org/10.1057/9780230599246_4

Leinhardt, G., Stainton, C., Virji, S. M., & Odoroff, E. (1994). Learning to Reason in History: Mindlessness to Mindfulness. In *Cognitive and Instructional Processes in History and the Social Sciences*. Routledge.

Lin, C.-C., Huang, A. Y. Q., & Yang, S. J. H. (2023). A Review of AI-Driven Conversational Chatbots Implementation Methodologies and Challenges (1999–2022). *Sustainability*, *15*(5), Article 5. https://doi.org/10.3390/su15054012

Liu, T.-C. (2022). Examining the Links Between Heritage Use and Identity Building Through

a Narrative Approach: A Case Study of Dadaocheng, Taipei City, in Taiwan. *Journal of Heritage Management*, *7*(1), 22–36. https://doi.org/10.1177/24559296211025178

Matsangou, E. (2018). *How Fortnite became the most successful free-to-play game ever*. https://www.theneweconomy.com/business/how-fortnite-became-the-most-successful-free-to-play-game-ever

McCall, J. (2016). Teaching History With Digital Historical Games: An Introduction to the Field and Best Practices. *Simulation & Gaming*, *47*(4), 517–542. https://doi.org/10.1177/1046878116646693

Messaris, P., & Humphreys, L. (2007). *Digital Media: Transformations in Human Communication* (2nd edition). Peter Lang Publishing Inc. New York.

Nachtigall, V., Yek, S., Lewers, E., Brunnenberg, C., & Rummel, N. (2022). Fostering cognitive strategies for learning with 360° videos in history education contexts. *Unterrichtswissenschaft*, *50*(4), 615–638. https://doi.org/10.1007/s42010-022-00154-x

Perfetti, C. A., Britt, M. A., & Georgi, M. C. (1995). *Text-based Learning and Reasoning: Studies in History*. Routledge. https://doi.org/10.4324/9780203053065

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751–783. https://doi.org/10.1037/0022-3514.90.5.751

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist*, *50*(4), 258–283. https://doi.org/10.1080/00461520.2015.1122533

Ropeik, R. (2014, October 2). *Game design*. Art Museum Teaching. https://artmuseumteaching.com/tag/game-design/

Schell, J. (2014). *The Art of Game Design: A book of lenses* (2nd rev ed. edition). Taylor & Francis.

Schreiber, W., Körber, A., Borries, B., Krammer, R., Leutner-Ramme, S., Mebus, S., Schöner, A., & Ziegler, B. (2006). *Historisches Denken: Ein Kompetenz-Strukturmodell [A competence structure model of historical thinking]*.

Squire, K., Gee, J. P., & Jenkins, H. (2011). *Video Games and Learning: Teaching and Participatory Culture in the Digital Age* (Illustrated edition). Teachers' College Press.

Sugiyama, M. (2001). Narrative Theory and Function: Why Evolution Matters. *Philosophy and Literature*, *25*, 233–250. https://doi.org/10.1353/phl.2001.0035

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

VanSledright, B. A., & Frankes, L. (2000). Concept- and Strategic-Knowledge Development in Historical Study: A Comparative Exploration in Two Fourth-Grade Classrooms. *Cognition and Instruction*, *18*(2), 239–283.

Viccari, A. (2016). *A proposal of an archaeological and heritage Role-Playing Video-game of the ancient Roman city of York (Eboracum).* [Unpublished Master's thesis]. University of York.

Wessell, A. (2017). "We Will Show the Country": Bringing History to Life. *M/C Journal*, *20*(5), Article 5. https://doi.org/10.5204/mcj.1299

Wineburg, S. S. (2001). *Historical Thinking and Other Unnatural Acts: Charting the Future of Teaching the Past*. Temple University Press,U.S.

# 3.2

# Paper 2

Viccari, A. & Bardach, L. (2025). Designing Limes Text: Adapting an educational video game into a text-based format for controlled experimental comparisons. *Proceedings of the European Conference on Games Based Learning, 19*(2), 855-863. https://doi.org/10.34190/ecgbl.19.2.3972

**Abstract**

Comparing educational video games and text-based learning presents a significant challenge when trying to ensure a fair comparison between the two formats. Their distinct natures, particularly in terms of engagement and interactivity, introduce uncontrolled variables that are often not directly comparable between a video game and a text. Many prior studies rely on static text conditions that do not account for reader engagement, decision-making, or interaction patterns. This paper therefore proposes a structured game design approach to developing a text-based application that enables a more precise and controlled comparison with video game-based learning. *Limes* is a 2D educational video game set in Roman frontier Germany, designed to teach historical knowledge through exploration and interactive learning elements. The game features a structured narrative, Codex entries with historical information, and secondary missions, all contributing to the player's engagement and learning experience. The text version of *Limes* is not a game but replicates key gameplay characteristics in a structured text-based format, providing an alternative to traditional text conditions in experimental research. Unlike passive reading conditions, this Text-app is designed to be interactive and engaging in the same ways as its video game counterpart. Moreover, like the video game, the Text-app tracks player behaviour, such as time spent reading, engagement with Codex entries, and whether secondary mission texts have been read, using a five-star rating system mirroring the game's assessment mechanics. The present paper discusses the development of the *Limes* Text-app and presents a game design framework for adapting educational games into experimentally viable text-based counterparts. The contribution is particularly relevant to research on video game effects in education adopting experimental designs. However, beyond its experimental utility, the newly developed text-app offers promising opportunities for other research areas, and we discuss how it can, for example, be used to assess and explore digital traces of learner characteristics and behaviours. To conclude, by ensuring structural comparability and capturing interaction data across formats, the *Limes* Text-app addresses a key methodological challenge in video game-based learning research: the difficulty of creating valid text-based control conditions that enable meaningful comparisons. While *Limes* has been developed in the context of history education, the impact is much broader as our design framework is adaptable and could be extended to other domains and narrative-based games.

*Keywords:* Game design, text-based learning, educational gaming, history education, experimental methodology

## 1. Introduction

The application of digital technologies in education and cultural heritage has proven to be a promising avenue for engaging diverse audiences through immersive, dynamic, and personalized 1. Introduction

Over the past two decades, educational video games have gained increasing recognition for their potential to enhance learning outcomes, particularly by fostering engagement, motivation, and deep cognitive processing (Clark et al., 2016; Ren et al., 2024; Wouters et al., 2013). This growing body of research has resulted in numerous experimental comparisons of video game-based learning with more traditional learning methods, most commonly text-based formats. However, these comparisons often suffer of methodological limitations related to structural differences between formats, particularly the use of passive text conditions (e.g., slideshows or static PDFs) that lack interactivity, decision-making, or behavioural tracking—an issue repeatedly flagged in systematic reviews (Gui et al., 2023; Riopel et al., 2019; Wouters et al., 2013). These limitations can undermine experimental validity, as they make it difficult to determine whether observed effects and differences between formats are due to the format itself or to differences in interactivity and design features (All et al., 2014).

For example, Gutierrez et al. (2014) compared a video game aiming to promote perspective-taking with a text version of the same story. Even though the content was identical, the two conditions differed fundamentally in structure, and only the game offered interactivity and decision-making. The findings of the study revealed that the game condition led to reductions in implicit bias; nonetheless, given the differences between the conditions, it was impossible to disentangle whether the format (game versus text) or other features (interactivity and decision-making, which was only possible in the game condition) drove the effects. This also highlights how even content-equivalent text formats can fail to replicate the mechanisms through which games operate. More recently, Lee and Chen (2023) found that while both game and text interventions had a positive effect on intergroup attitudes, only the game condition allowed for tracking of user behaviour—a key limitation in the passive text condition, which offered no insight into decision-making, time-on-task, or interaction patterns.

Gui et al. (2023), drawing on Mayer´s (2015) framework, classify most digital game-based learning (DGBL) studies into two main types: media comparison studies—which contrast games with traditional formats such as text or lectures—and value-added studies—which compare different versions of a game to isolate the effects of specific design elements (e.g., feedback, adaptation). While Gui et al. (2021) do not focus directly on control group design,

they emphasise that meaningful comparisons require attention to how learners interact with design elements rather than focusing solely on the delivery format. This perspective reinforces the need for structurally comparable experimental conditions that can capture user behaviour and engagement, even in non-game formats.

The methodological issue of comparing game versus text became evident during the design of our experimental studies, which involved *Limes*, a 2D educational adventure game set in ancient Roman times in Germany (see Viccari et al., 2024, for a description of the game development). The initial idea for the control group was to read a simple PDF version of the game´s story. However, it quickly became apparent that the limitations of this format did not allow a valid comparison with the video game. Also, key behavioural measurements that can be accessed during the gameplay of *Limes* cannot be equally measured with a standard PDF file. For example, crucial points for our research include tracking and measuring the selection of optional Codex entries, exploratory decisions—specifically, the accomplishment of the secondary missions—and a five-star rating system for level accomplishment, which are central to the *Limes* gameplay experience.

In response to these issues, we developed a structured, interactive text-based application built in Unity that replicates the game's narrative structure, meant to track and measure the key game mechanics mentioned above. This Text-app represents a novel approach to control group design in video game-based learning research, one that ensures a functional comparison between a text and a video game, allowing researchers to hold specific key mechanisms constant across the two formats.

This design paper presents the process and the decision-making behind the development of the *Limes* Text-app and proposes a framework for converting educational video games into interactive text-based counterparts. By addressing the limitations of passive text control conditions, this paper offers a methodological contribution to the field of experimental educational research. Through this framework, we aim to provide a foundation for more valid and controlled comparisons between video games and text-based learning formats, ensuring that differences in outcomes can be more confidently attributed to format effects.

## 2. How to Convert a Video Game into Text

### 2.1 The Game

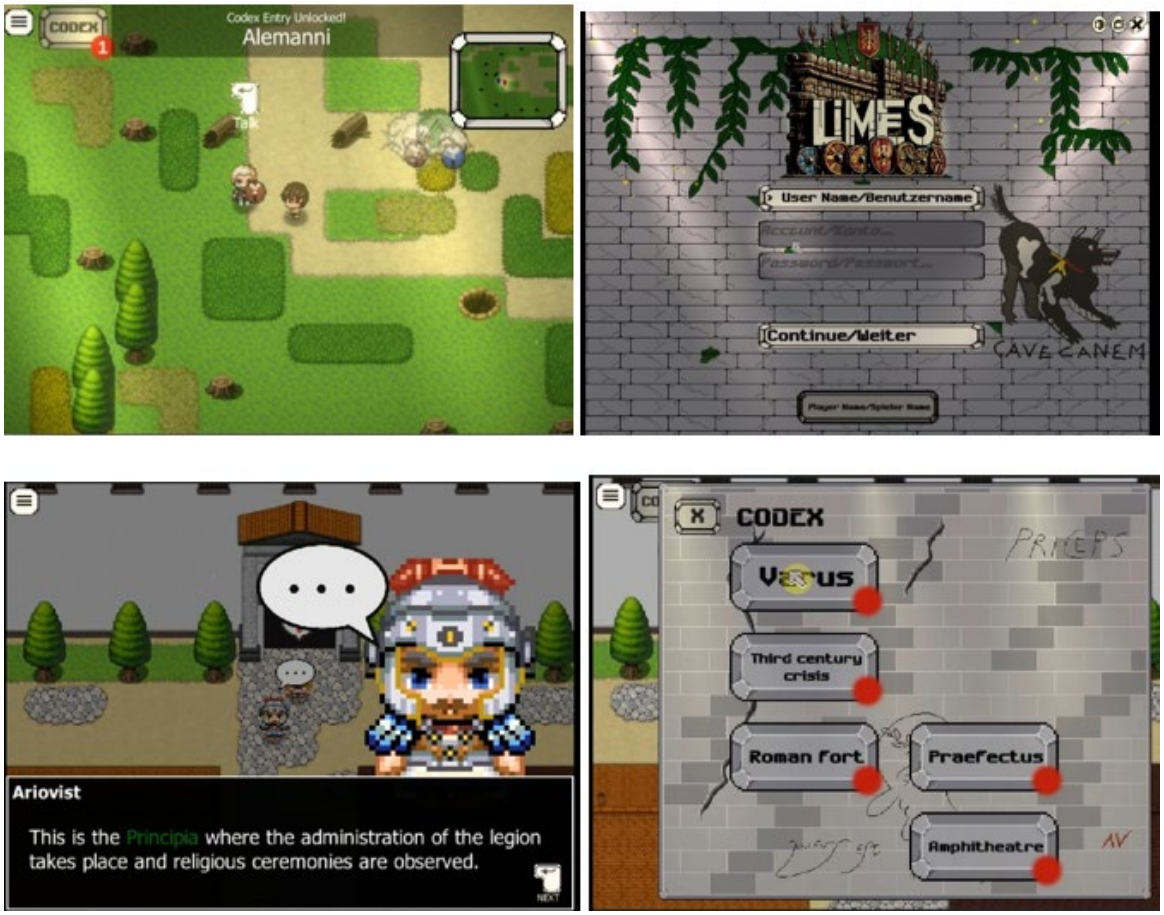
*Limes*[3] is a 2D educational video game developed in Unity and designed for 5th and 6th-grade students in Germany, focusing on Roman history. The video game story is set in third-century southern Germany; the game follows the journey of an Alemanni child and a Roman centurion named Ariovist in a period where the tensions between the Roman Empire and the Alemanni, a prominent Germanic tribe, are rising. The narrative is structured across five levels; each level contains a main and a secondary mission that rewards players through a five-star rating system. Players can engage in varied game mechanics, including stealth, exploration, dialogue choices, horse racing, puzzle-solving, and rock-throwing.

Players can obtain a considerable amount of historical information from the dialogues presented during the main and secondary missions. However, an essential pedagogical feature is the Codex system, which allows players to access historical content embedded in the game world (Figure 1). These entries are unlocked through interaction with the Non-Playable-Characters (NPCs) (Figure 1), providing deeper context and more historical information about the Roman period. These behaviours are also trackable, providing researchers with crucial behavioural information, such as interactions with the codex entries, the accomplishment of secondary missions, and time-on-task tracking. The video game has been designed with input from history education, psychology, and game-based learning principles; *Limes* aims to foster historical knowledge, positive diversity beliefs, and motivation through a narrative-driven, interactive experience (Viccari et al., 2024).

[3] The *Limes* video game can be accessed at: **https://ancient-verse-research.itch.io/limes** (Windows and Mac versions available).

**Figure 1**

*(Top-left) Gameplay screenshot in Limes video game. (Top-right) Login screen of the Limes video game. (Bottom-left) An example of a Codex entry being unlocked during gameplay. (Bottom-right) The Codex menu shows unlocked entries during gameplay.*




## 2.2 Identifying Core Game Features for Adaptation

While educational video games may include various immersive and interactive elements, not all are essential—or feasible—for replication in a text-based format. Identifying the suitable mechanics in the *Limes* video game was a crucial stage in creating a comparable text-based control condition. It was decided to select features that support the intended learning outcomes and represent behavioural data relevant to the experimental study. The following elements were selected as core to the adaptation process.

**Codex entries** are a central mechanic in *Limes* (Figure 2), allowing players to read or hear historical information connected to in-game dialogues and explore Roman history in greater depth. The content was adapted for 5th and 6th-grade students, covering key topics from the school curriculum. These entries are optional and are unlocked through exploration or dialogue. Their inclusion is pedagogically significant, as they support factual knowledge

acquisition and allow learners to engage with supplementary material at their own pace. As Mayer (2015) notes, narrative context and thematic relevance can enhance learner motivation and engagement. According to relevance-based reading models, learners are more inclined to engage with optional content if they find it meaningful (McCrudden et al., 2010; McCrudden & Schraw, 2007). For this reason, the text version needed to replicate both the availability of these entries and the ability to track interactions, specifically, how often they were accessed and how much time was spent on them.

**Figure 2**

*Codex entry example during a gameplay in Limes video game.*


**Secondary missions** function as exploratory and additional side quests, allowing players to interact with additional narrative material, deepen their understanding of the game world, and access extra game mechanics, such as horse racing. These missions are optional, reflecting the player's autonomy in engaging with extended content. From a design perspective, they represent an opportunity to allow for learner agency, autonomy, and interest-driven behaviour, thus contributing to their self-determined motivation (Dennie, 2012; Proulx et al., 2017; Ryan & Deci, 2000). In adapting the game into a text format, it was important to preserve this element of choice, allowing readers to decide whether to access the additional material while also enabling that choice to be tracked.

**Time-on-task tracking** was another essential feature that needed to be included in the text adaptation of *Limes*. In the original game, the Unity code records the amount of time spent on several sections and transfers that information to the Mongo Database —such sections are the main storyline, secondary missions, and Codex entries. Tracking time spent on different sections provides insight into learners' attention, engagement, and navigation behaviour within the *Limes* environment. Therefore, incorporating equivalent tracking in the text version was necessary to allow for behavioural comparison across conditions. This approach aligns with

recent research advocating for objective behavioural tracking in digital environments, as it provides more accurate insights into user engagement than self-reported measures (Johannes et al., 2021).

The game's **five-star rating system** is an integrated assessment mechanism that evaluates performance based on exploration, completion of secondary missions, and Codex usage. This system is pedagogically relevant because it provides immediate, performance-based feedback, which has been shown to foster learners' satisfaction of their need for competence, and their motivation (Bardach & Murayama, 2025; Sailer et al., 2017). Recreating this rating system in the Text-app was important to allow participants in both conditions to receive similar performance feedback and to maintain motivational comparability across formats.

Together, these four features—Codex interaction, secondary mission engagement, time-on-task in the different sections of the game, and the five-star feedback system—were identified as central to the pedagogical structure of *Limes* and critical for experimental comparability. Their inclusion in the text version was not an attempt to gamify the reading experience but rather to create a structurally aligned learning environment that would allow for meaningful behavioural tracking and outcome comparison in a controlled study.

### 2.3 Framework for Game-to-Text Adaptation

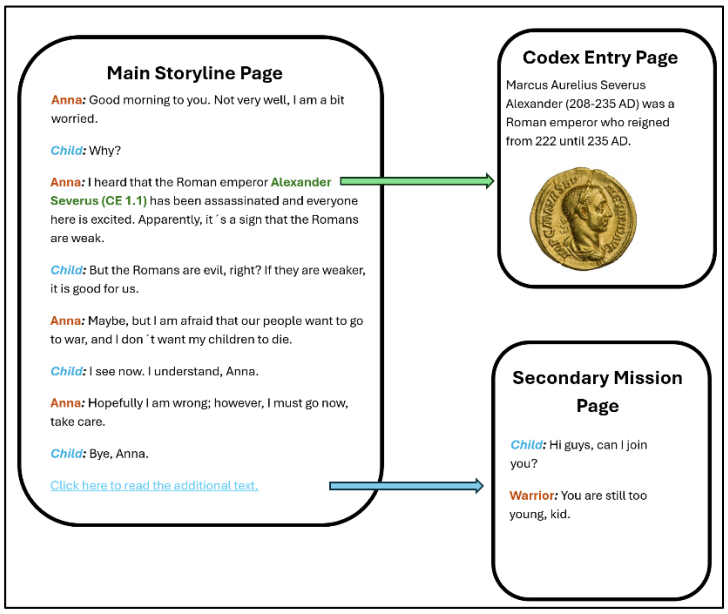
Challenges discussed earlier in the introduction highlighted the need for a control condition that goes beyond a static PDF or linear text, offering greater interactivity and measurability. Considering the experience gained while developing the *Limes* video game, we opted for a structured text application[4] using Unity. This allowed for precise segmentation of content and the implementation of custom tracking systems mirroring the logic of the game (Figure 3). The text was divided into three main types of content:

1. Main storyline,
2. Secondary missions
3. Codex entries

[4] *Limes* Text-app can be accessed at: **https://ancient-verse-research.itch.io/limes-version-2** (Windows and Mac versions available).

**Figure 3**

*The diagram provides a simplified illustration of the Unity-based structure, showing how the main storyline connects to optional Codex entries and secondary missions, each displayed in separate text panels.*




Each section was implemented as an interactive module with automated time tracking and click logging.

The **main storyline** served as the structural backbone of the app and was set as the default view that loaded when the user launched the application. It was implemented as a scrollable text panel within Unity's system, very similar to a regular PDF document, with time tracking initiated automatically upon scene load. Since all study participants are required to read the main story, this section provided the baseline to which optional interactions, like secondary missions and Codex entries, could be added.

**Secondary missions**, which were optional side quests in the game, were similarly adapted as clickable links embedded in the main narrative (Figure 4). These links appeared as light blue text at relevant decision points of the story, such as "Click here to read the additional text."

**Figure 4**

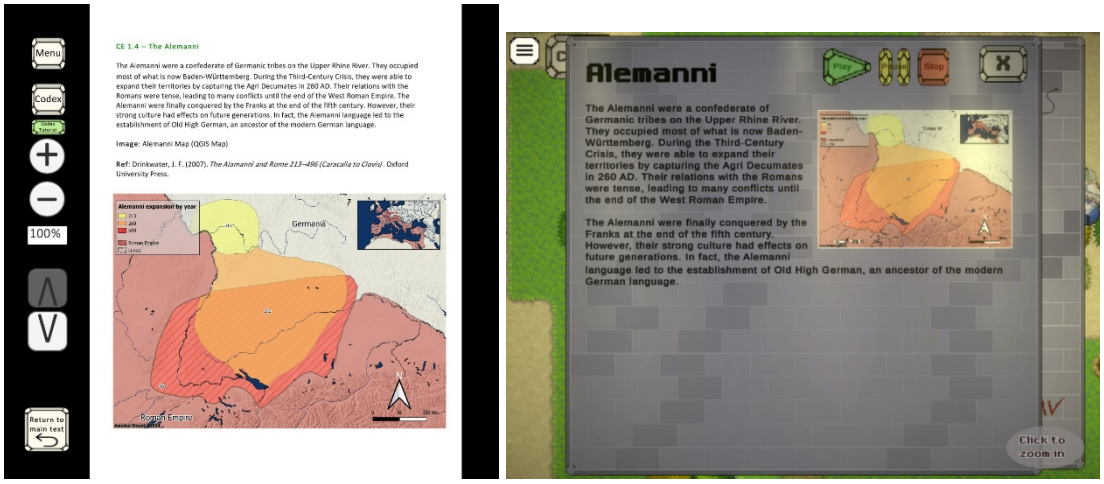
*(Left) additional text link in the Text-app, (right) secondary mission choices during the gameplay of Limes video game.*


When the link is clicked, the app records the interaction and the time spent on the additional text page(s). This enables precise measurement of engagement with optional content—an essential feature to support autonomy and potential indicator of learners' interest.

**Codex entries**, another optional learning mechanic in the game, were embedded directly into the main narrative using the same green-highlighted format as in the original game (Figure 1). When clicked, each Codex entry opened in a dedicated window (Figure 5). As with the game counterpart, the system logged the time spent reading each entry and whether it had been accessed. All interaction data was stored in a Mongo Database, identical to the architecture used for logging gameplay behavioural data in *Limes*.

**Figure 5**

*(Left) Codex entry example in the Text-app. (Right) Codex entry example in Limes video game.*




The **five-star rating system** was also implemented in the text version. This allows users to see their performance for each section of the text, equivalent to the levels in the game version.

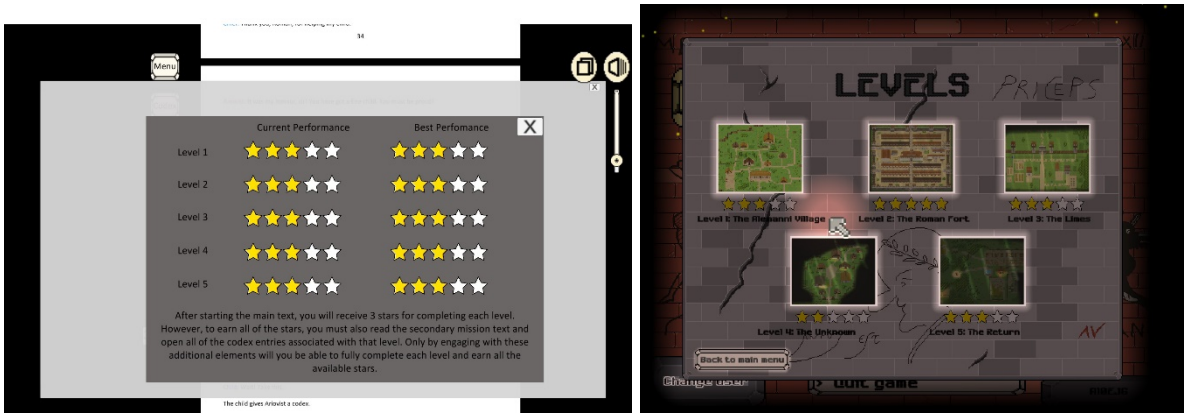
The performance is based on these actions:

- Clicking all Codex entries available in a text section.
- Access the additional text link corresponding to the secondary mission in the gameplay.

As in the game version, the Text-app requires users to complete the main storyline before receiving their star rating, which is displayed in a similar format to maintain consistency across the two media (Figure 6).

**Figure 6**

*(Left) Five stars awarding system in the Text-app. (Right) Five stars awarding system in Limes video game.*




Another important aspect replicated in the Text-app is full access to the Codex entries once the main text has been completed. As in the game, all Codex entries become accessible at this point; however, accessing them after completing the main content does not affect the five-star rating—unless the reader revisits the corresponding text sections, just as players would need to replay levels in the game.

The Text-app preserves the same narrative and allows users to make similar decision options by clicking on additional text or the codex entries, while also enabling precise tracking of these behaviours. This approach is meant to establish a structurally aligned control condition, ensuring that any differences in learning outcomes are more likely connected to the format itself (game vs. text), rather than to variations in content structure, engagement features, or users' behaviours.

### 2.4 Early Testing and Feedback

As with the original *Limes* video game, the text-based application was tested with experts during development to gather formative feedback on content clarity and possible

technical problems. These reviewers included educational researchers and historians familiar with the project's goals, as well as individuals experienced with digital learning tools.

Since the app was developed in both German and English, preliminary testing was conducted in both languages. However, special attention was given to the German version, as it was the version intended for use our first experiments. This included reviewing the linguistic accuracy of the texts and checking for fluency and user comfort when navigating the app's interface.

Feedback led to two main refinements. First, some dialogue passages and transitions were restructured slightly to better fit a linear reading experience, particularly in the German version. Second, the initial scroll system was unintuitive and slow, presenting several issues when combined with the navigation arrows. This resulted in delays and imprecise navigation through the text. To address this, the underlying code was adjusted to ensure smoother scrolling and to make the app interface more user-friendly. These changes significantly improved navigation fluidity across the main story, Codex entries, and secondary missions while preserving the behaviour-tracking structure required for the experiment.

A crucial aspect of the testing phase was verifying the interaction between the Text-app and the Mongo Database, where behavioural data, such as Codex clicks and time-on-section, is stored. The tests confirmed that the data was recorded accurately and reliably. While the video game and Text-app use separate MongoDB databases with different field names and structures, the underlying logic and data types are aligned. These differences are the results of the independent Unity development of the two applications. Both systems record the same core variables: time spent (in seconds), Codex interactions (number of clicks), star achievements per level, and secondary mission completion, see below:

| Mongo Database – Video game | Mongo Database – Text-app |
|---|---|
| **totalTimeLevel1:** seconds | timeSpentMainText: seconds |
| **totalTimeLevel2:** seconds | timeSpentCodex: seconds |
| **totalTimeLevel3:** seconds | timeSpentSecondaryMission: seconds |
| **totalTimeLevel4:** seconds | codexEntryClicks: number of clicks |
| **totalTimeLevel5:** seconds | totalActiveTimeMainTextSecondaryMission: seconds |
| **totalTimeMainSec:** seconds | level_1_stars: number of stars |
| **totalTimeCodexMenuOpen:** seconds | level_2_stars: number of stars |
| | level_3_stars: number of stars |

| | |
|---|---|
| **codexButtonClicked:** number of clicks<br>**starsLevel1:** number of stars<br>**starsLevel2:** number of stars<br>**starsLevel3:** number of stars<br>**starsLevel4:** number of stars<br>**starsLevel5:** number of stars<br>**CompletedSecondaryMissions:** Object 1 (Mission Achieved) | level_4_stars: number of stars<br>level_5_stars: number of stars<br>level_[level number]_secondary_star_awarded: Mission Achieved – true or false |

In the game, mission completion is logged as an object per level; in the Text-app, the same is recorded as Boolean values (true of false). Time tracking of the secondary missions is split in the Text-app for structural reasons but later merged—mirroring the unified level time in the game.

### 2.5 Limitations and Future Directions

While this paper does not report empirical results from experiments, its contribution lies in presenting a structured and trackable text-based format that enables more valid comparisons between video games and texts in experimental research. This constitutes, to the best of our understanding, the first attempt of its kind. Our paper provides a practical framework that other researchers can adapt or build upon when designing their own structurally aligned control conditions. While tailored to a narrative-driven history game, the framework may be transferable to other domains that rely on exploratory learning or decision-making, such as science simulations or cultural education. Its modular structure and tracking logic are adaptable to any learning context involving optional content, time-on-task, and user-driven exploration. Future work could explore how this framework performs when adapted to other domains (e.g., environmental science, language learning, or civic education), where narrative and autonomy also play key roles. Overall, our approach has the potential to inspire similar projects in other domains by showing how behavioural tracking and narrative structure can be replicated in non-game formats. Future work will evaluate the empirical utility of this framework and examine how structural alignment, and behavioural tracking can support robust comparisons across different educational formats. In our own upcoming experimental studies comparing the *Limes* video game and the *Limes* Text-app, we aim to gain insights into the pedagogical effectiveness of such a Text-app compared to its game counterpart, particularly in terms of learning outcomes (e.g., knowledge on the Roman Empire), motivation, engagement patterns, and intercultural

attitudes.

Beyond its applicability in experimental contexts, the newly developed Text-app opens up additional opportunities for future research. There is growing interest in educational and psychological research in modelling and assessing digital traces of learner characteristics and behaviours to test theories in action and inform practice. For example, researchers are exploring digital traces of educationally relevant personality traits and motivations, and use digital traces to predict outcome variables, such as test performance and knowledge attainment (e.g., Tlili et al., 2019; Matcha et al., 2020; Schumacher et al., 2025). Exploring whether different learner characteristics are reflected in behavioural patterns captured by the *Limes* Text-app and the video game could yield valuable insights for research and offer practical implications for adaptive educational technologies. In addition to using the behavioural indicators described in the present design paper, researchers could also create new indicators to more comprehensively test their theoretical assumptions in future research with *Limes*. For instance, a recent study found some indications that intellectual curiosity (as a personality trait) and domain-specific interest (as a motivational construct) manifested in different information-seeking behaviours: interested learners employed a more widespread search for information, whereas curious individuals engaged more deeply with the learning content by revisiting information more frequently (Schumacher et al., 2025). This distinction could be further investigated using *Limes*, for example, by analysing the revisiting of codex entries as a behavioural indicator.

## 3. Conclusion

This design paper introduced the development of a structured, interactive Text-app built as a control condition for experimental research comparing video game- and text-based learning formats in history education. The text application replicates the key pedagogical mechanics of the educational video game *Limes*—including time-on-task tracking, engagement with optional content, and a five-star feedback system—within a Unity-based environment that enables precise behavioural monitoring. By ensuring structural comparability and capturing interaction data across formats, the Text-app addresses a significant methodological challenge in video game-based learning research: the difficulty of creating valid text-based control conditions that allow meaningful comparisons. Even though the current project is tailored to Roman history and the *Limes* project, the framework is adaptable and could be extended to other domains and narrative-based games.

**References**

All, A., Castellar, E. P. N., & Looy, J. V. (2014). Measuring Effectiveness in Digital Game-Based Learning: A Methodological Review. *International Journal of Serious Games*, *1*(2), Article 2. https://doi.org/10.17083/ijsg.v1i2.18

Bardach, L., & Murayama, K. (2025). The role of rewards in motivation—Beyond dichotomies. *Learning and Instruction*, *96*, 102056. https://doi.org/10.1016/j.learninstruc.2024.102056

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital Games, Design, and Learning: A Systematic Review and Meta-Analysis. *Review of Educational Research*, *86*(1), 79–122. https://doi.org/10.3102/0034654315582065

Dennie, T. (2012). *Perception of autonomy and its effect on intrinsic motivation, immersion, and performance*. https://www.semanticscholar.org/paper/Perception-of-autonomy-and-its-effect-on-intrinsic-Dennie/d521427cd8253f5854b1bdccae8681d817f1617d

Gui, Y., Cai, Z., Yang, Y., Kong, L., Fan, X., & Tai, R. H. (2023). Effectiveness of digital educational game and game design in STEM learning: A meta-analytic review. *International Journal of STEM Education*, *10*(1), 36. https://doi.org/10.1186/s40594-023-00424-9

Gutierrez, B., Kaatz, A., Chu, S., Ramirez, D., Samson-Samuel, C., & Carnes, M. (2014). "Fair Play": A Videogame Designed to Address Implicit Race Bias Through Active Perspective Taking. *Games For Health Journal*, *3*(6), 371–378. https://doi.org/10.1089/g4h.2013.0071

Johannes, N., Vuorre, M., & Przybylski, A. K. (2021). Video game play is positively correlated with well-being. *Royal Society Open Science*, *8*(2), 202049. https://doi.org/10.1098/rsos.202049

Lee, Y.-H., & Chen, M. (2023). Intergroup contact with a virtual refugee: Reducing prejudice through a cooperative game. *Psychology of Popular Media*, *12*(4), 383–392. https://doi.org/10.1037/ppm0000433

Matcha, W., Gašević, D., Jovanović, J., Uzir, N. A., Oliver, C. W., Murray, A., & Gasevic, D. (2020). Analytics of learning strategies: The association with the personality traits. *LAK 2020 Conference Proceedings: International Learning Analytics & Knowledge Conference 2020*, 151–160. https://doi.org/10.1145/3375462.3375534

Mayer, R. E. (2015). On the Need for Research Evidence to Guide the Design of Computer Games for Learning. *Educational Psychologist*, *50*(4), 349–353. https://doi.org/10.1080/00461520.2015.1133307

McCrudden, M. T., Magliano, J. P., & Schraw, G. (2010). Exploring how relevance instructions affect personal reading intentions, reading goals and text processing: A mixed methods study. *Contemporary Educational Psychology*, *35*(4), 229–241. https://doi.org/10.1016/j.cedpsych.2009.12.001

McCrudden, M. T., & Schraw, G. (2007). Relevance and Goal-Focusing in Text Processing. *Educational Psychology Review*, *19*(2), 113–139. https://doi.org/10.1007/s10648-006-9010-7

Proulx, J.-N., Romero, M., & Arnab, S. (2017). Learning Mechanics and Game Mechanics Under the Perspective of Self-Determination Theory to Foster Motivation in Digital Game Based Learning. *Simulation & Gaming*, *48*(1), 81–97. https://doi.org/10.1177/1046878116674399

Ren, J., Xu, W., & Liu, Z. (2024). The Impact of Educational Games on Learning Outcomes: Evidence From a Meta-Analysis. *International Journal of Game-Based Learning (IJGBL)*, *14*(1), 1–25. https://doi.org/10.4018/IJGBL.336478

Riopel, M., Nenciovici ,Lucian, Potvin ,Patrice, Chastenay ,Pierre, Charland ,Patrick, Sarrasin ,Jérémie Blanchette, & and Masson, S. (2019). Impact of serious games on science learning achievement compared with more conventional instruction: An overview and a meta-analysis. *Studies in Science Education*, *55*(2), 169–214. https://doi.org/10.1080/03057267.2019.1722420

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior*, *69*, 371–380. https://doi.org/10.1016/j.chb.2016.12.033

Schumacher, A., Kammerer, Y., Scharinger, C., Gottschling, S., Hübner, N., Tibus, M., Kasneci, E., Appel, T., Gerjets, P., & Bardach, L. (2025). How do intellectually curious

and interested people learn and attain knowledge? A focus on behavioral traces of information seeking. *European Journal of Personality*, 08902070241309124. https://doi.org/10.1177/08902070241309124

Tlili, A., Denden, M., Essalmi, F., Jemni, M., Kinshuk, Chen, N.-S., & Huang, R. (2019). Does Providing a Personalized Educational Game Based on Personality Matter? A Case Study. *IEEE Access*, *7*, 119566–119575. https://doi.org/10.1109/ACCESS.2019.2936384

Viccari, A., Göllner, R., Hahn, J.-U., & Bardach, L. (2024). Limes (the Roman Frontier): Developing a Video Game for History Learning. *Proceedings of the 18th European Conference on Games Based Learning*, *18*, Article 1. https://doi.org/10.34190/ecgbl.18.1.2833

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, *105*(2), 249–265. https://doi.org/10.1037/a0031311

.

# 4

# EXPERIMENTAL STUDIES

# 4.1

## Paper 3

Viccari, A., Kienatz, A., Eitel, A., Hahn, J-U., Göllner, R., Lindner, M., & Bardach, L. (2026) Exploring Roman history through videogame and text: Effects of format and instructional framing. *Manuscript under review.*

Preprint can be accessed at https://doi.org/10.31234/osf.io/fa28t_v1

**Abstract**

This study investigated effects of content format and instructional framing on knowledge attainment, situational interest, and academic emotions in digital history learning. In a 2 (format: game versus text) × 2 (instruction: learning versus fun) between-subjects design, 143 university students engaged with either a historically accurate video game on the Roman Empire or a structurally aligned interactive text application, each paired with instructions to focus on learning or fun. Both formats presented the same historical content through main narratives, optional missions, and supplemental historical information in the form of codex entries. The findings revealed that participants in the game condition reported higher enjoyment and greater triggered situational interest, indicating that the game enhances the positive affective experience of engaging with historical content and sparks learners' interest. However, our hypothesis that knowledge test performance would be higher in the game condition was not supported. Instead, knowledge test scores were higher in the text condition. No statistically significant effects emerged for instructional framing or for the interaction between format and framing on the outcomes. Additional analyses of behavioural data showed that text participants engaged more extensively with codex entries, and that codex engagement strongly predicted knowledge outcomes across both formats. This suggests that knowledge outcomes depend on the visibility and integration of supplemental materials, and learners' engagement with them. Overall, this study contributes to research on digital history learning by offering a controlled comparison of identical content delivered through game and text and shedding light on the role of instructional framing.

*Keywords:* Video game, History learning, Roman history, Academic outcomes, Motivation, Academic emotions, Format comparison

## Introduction

Over the past decades, digital games have gained increasing recognition as meaningful tools in education (Clark et al., 2016). Their value lies not only in their motivational appeal, but also in their capacity to engage learners with complex subjects, such as history, through interactive, narrative-driven, and exploration-based environments (Gee, 2003; McCall, 2016). While traditional media such as textbooks and documentaries also aim to present the past, they typically lack the interactivity and decision-making opportunities that allow learners to actively engage with historical complexity. In the context of history learning, these qualities of digital games offer new ways to embed knowledge into immersive narratives, allowing learners to navigate historical environments, confront dilemmas, and explore different perspectives (Chapman, 2016; Gee, 2003; McCall, 2016).

Such game-based engagement offers an alternative to traditional textbooks by making historical content more tangible and experiential, aligned with contemporary pedagogical goals. Rather than focusing solely on the memorization of facts and dates, current models of history education emphasize historical reasoning, critical reflection, and the development of historical consciousness as key competencies (van Boxtel & van Drie, 2018; Wineburg, 2001). Well-designed video games can support these aims by immersing learners in dynamic historical settings that demand interpretation, decision-making, and reflection—skills central to historical reasoning. This can foster both knowledge acquisition and reflective engagement with the past (McCall, 2016, 2023; Spring, 2015).

Additionally, digital games may help address a common challenge in history learning: learners often interpret historical content through presentist lenses, which can lead to oversimplified or anachronistic conclusions (Akkerman et al., 2009). Interactive, decision-driven narratives can help mitigate these issues by immersing learners in historically based-scenarios that encourage temporally contextualized thinking (Chapman, 2016). Recent scholarship highlights the potential of educational technologies—including games, virtual environments, and mobile platforms—to enrich history learning through multi-sensory and inquiry-based experiences (Clyde et al., 2012; Crymble, 2021; Kusuma et al., 2021). This is further evidenced in museum contexts, where digital storytelling has been employed to create immersive and educational experiences for visitors (e.g., Rosli et al., 2022).

Yet, despite this growing interest, empirical research on the use of history-themed games remains limited. Some purpose-built games have demonstrated promising outcomes, showing that historically grounded digital environments may foster empathy, contextual reasoning, and

conceptual understanding (Schaper et al., 2018; Spring, 2015). At the same time, widely used commercial titles such as the Civilization series (Firaxis Games, 1991–present) or the Assassin's Creed series (Ubisoft, 2007–present) have been praised for sparking interest in historical topics. Because such games often prioritize entertainment and gameplay over educational depth (McCall, 2016), however, they frequently present simplified or distorted narratives. Consequently, they risk undermining meaningful historical understanding when used uncritically in learning contexts (Robison, 2013). These pedagogical challenges highlight the need for historically accurate, empirically validated educational games that not only engage learners but also support cognitive outcomes. Achieving this requires systematic research that integrates game-based experiences with history learning, and compares the effect of games to other widely used formats (e.g., text-based history learning).

It is also essential to consider how instructional framing shapes engagement and learning in digital history learning environments. Games can motivate through interactivity and immersion: however, how instructional goals are presented may influence knowledge attainment, especially in adventure-style formats where strong narrative and entertainment features may risk overshadowing pedagogical aims. The present study therefore aimed to investigate the effects of content format and instructional framing on academic outcomes (knowledge test performance, situational interest, and academic emotions) in the context of digital history learning. Participants were randomly assigned to one of four conditions in a 2 (format: video game vs. text) × 2 (instruction: learning vs. fun) between-subjects design and engaged with *Limes*, a video game and its structurally aligned text version that focus on Roman history. Overall, this study contributes to the growing body of experimental research on digital learning games and digital history learning by offering a controlled comparison of identical content delivered through game and text formats, while also shedding light on the role of instructional framing.

**Effects of video games on academic outcomes**

A primary goal of educational games is to foster knowledge acquisition (Gee, 2003; McCall, 2023). Several meta-analyses have shown that digital games can enhance learning outcomes across different domains (e.g., Clark et al., 2016). In the context of history education, well-designed games may support learners in navigating historical narratives, interpreting causality, and evaluating multiple perspectives. McCall (2023) highlights how simulation-based historical games can promote historical thinking by immersing students in active exploration, decision-making, and temporally contextualized reasoning. Similarly, Wainwright

(2014) discusses that complex game environments can make abstract historical theories, such as debates around structure and agency, more accessible by enabling learners to engage with them through gameplay. A recent example is *The Darkest Files* (Paintbucket Games, 2025), a German historical adventure game that invites players to investigate real war crimes, thereby foregrounding tensions between individual agency and systemic structures through interactive storytelling.

In addition to supporting knowledge acquisition, educational games are assumed to enhance motivational factors such as interest (e.g., Bardach & Murayama, 2025). This may be particularly relevant in history, where learners often find it difficult to connect the content to their own lives (Barton & Levstik, 2004). The subject is often perceived as abstract, temporally distant, or disconnected from students' everyday experiences. These perceptions can hinder learners' motivation and engagement, making it more difficult to achieve deep understanding (Barton & Levstik, 2004; Stoel et al., 2015). Educational games embed historical content in interactive narratives and decision-driven gameplay. These features may cause situational interest for the learning content, an affective-cognitive state triggered by features of the learning environment such as novelty, personal relevance, and interactivity (Hidi & Renninger, 2006; Linnenbrink-Garcia et al., 2010). Drawing on the framework of Hidi and Renninger (2006), such immersive environments may trigger situational interest through emotional and role-based engagement, and maintain situational interest over time (Linnenbrink-Garcia et al., 2010). For example, Rotgans and Schmidt (2017) argued that instructional strategies, such as problem-based learning and relevance-focused framing, can successfully trigger situational interest and potentially help maintain it. Similarly, empirical work has demonstrated that tasks designed to promote causal reasoning, narrative construction, and epistemological reflection can support deeper engagement with historical material (Stoel et al., 2015). According to Hidi and Renninger (2006), these task features play a key role in both triggering and sustaining situational interest, thereby enhancing learners' attention, effort, and depth of understanding. Given that novelty, decision-making, and agency are known to spark situational interest (Hidi & Renninger, 2006; Murayama et al., 2019), game-based formats may provide a particularly motivationally supportive context for learning historical content. However, while some commercial games—such as *Assassin's Creed*—offer highly immersive historical environments that may trigger situational interest through visual realism and action-based storytelling, they are typically not designed with educational objectives in mind (Chapman, 2016).

Furthermore, an important aspect of engagement in learning involves students' academic emotion, such as enjoyment, boredom, frustration, and pride, that are directly tied to academic tasks (Pekrun, 2006). Among these, enjoyment and boredom are particularly important because they sit at opposite ends of the emotional spectrum, frequently occur in digital learning environments and are relevant to important student outcomes (Sakaki et al., 2024; Tze et al., 2016). While boredom is associated with inattention and disengagement, enjoyment supports persistence, curiosity, and deep learning strategies (Pekrun et al., 2002).

According to Control-Value Theory (Pekrun, 2006) academic emotions stem from learners' perceived control and value. In game-based learning, features such as autonomy-supportive mechanics, narrative immersion, and historical authenticity can enhance perceived control and reinforce value. When both control and value are appraised positively, enjoyment arises as a central activity-related emotion, fostering deeper engagement with complex historical narratives. By contrast, boredom typically emerges when value is low or control is undermined (e.g., Pekrun et al., 2002; Tze et al., 2016). El-shara (2015) emphasizes that adaptive emotions—especially when guided by thoughtful pedagogy—open space for deeper reflection and meaning-making in history education. This supports the broader view that emotions are central—not supplementary—to how students make sense of the past. Accordingly, academic emotions become important outcomes to explore in the context of digital game-based history learning.

**Comparing video games and text-based formats**

To investigate the effectiveness of video game-based learning, video games are commonly compared to text formats. From a theoretical standpoint, one might expect video game-based formats to outperform text in promoting academic outcomes due to their inherent characteristics, as outlined in the previous section ("Effects of video games on educational outcomes"). Also, games integrate multiple forms of representation for learning content (e.g., visual, textual, spatial), and cognitive theories such as the cognitive theory of multimedia learning suggest that dual-channel processing enhances memory and understanding (R. E. Mayer, 2014). Thus, it seems reasonable to assume that, as compared to a text-based format, video game-based history learning may enhance knowledge attainment, interest, and positive academic emotions (enjoyment), while reducing negative academic emotions (boredom) (de Ribaupierre et al., 2014; Ortiz, 2014; Ricci et al., 1996; Stiller & Schworm, 2019).

Nevertheless, with respect to effects on knowledge attainment, it has been proposed that

text-based formats may also hold unique benefits. From the perspective of the Focus-Support-Reward-Squares model (Eitel et al., 2025), text-based formats may better support students focus attention on essential content (i.e., active goal-directed processing) by providing a more streamlined presentation that limits non-essential interactions (for studies that find better learning outcomes for non-video game-based formats, see (Adams et al., 2012; Pilegard & Mayer, 2016), for studies finding non-significant differences, see Adams et al., 2012; Brom et al., 2019; Kienitz et al., 2025).

Importantly, scholars have increasingly emphasized that meaningful format comparisons require structural equivalence—not just content parity, but alignment in terms of interactivity, pacing, and feedback mechanisms (All et al., 2014; Gui et al., 2023). Without such alignment, claims about media-specific effects remain inconclusive. Despite these concerns, few studies have implemented tightly controlled comparisons where games and text formats offer comparable opportunities for engagement and learning. This methodological gap, particularly the challenges in ensuring comparable measures and designs between game-based and non-game learning conditions, has been identified as a persistent issue in the literature (Riopel et al., 2019). More rigorous experimental designs that control for key structural features are needed to clarify how medium-specific factors influence outcomes such as situational interest, knowledge acquisition, and academic emotions. The present study addresses this need by systematically aligning the structural features of a game and its text-based counterpart to enable a more valid comparison of their educational impact.

**Different instructions in digital learning games**

Instructional framing plays a role in shaping how learners engage with educational content. The goal-focusing model by McCrudden and Schraw (2007) suggests that emphasizing the relevance of specific goals helps learners direct attention to task-relevant information and improve recall, comprehension, and learning efficiency. While this effect has been confirmed across various domains (e.g., McCrudden et al., 2010), it may be especially important in digital game-based learning, where learners face the dual demands of playing and learning. While features of video games can enhance motivation, they may also lead learners to focus on entertainment rather than educational goals—especially when guidance is minimal. Simple but targeted instructional frames—such as telling learners to ‘play’ or ‘learn’—may significantly influence how they interpret the task, prioritize engagement, and allocate cognitive resources. Even minimal instructions can shape students’ metacognitive strategies and depth of processing, particularly in exploratory environments like digital games. In games, such framing

may play a more decisive role than in traditional text formats, where fewer distractions and clearer reading conventions guide attention (i.e., interaction between format and instruction). For instance, Erhel and Jamet (2013) found that framing a digital game as a learning activity improved knowledge acquisition. To conclude, learners do not necessarily know how to use games productively, and instructional framing focused on learning can help scaffold attention and behaviour toward achieving learning goals (R. E. Mayer, 2014; McCall, 2023). This issue may be especially relevant for history games, where students often engage passively with historical narratives unless explicitly guided to think critically (McCall, 2023; Wineburg, 2001).

**The present study**

The present study aimed to investigate the effects of content format and instructional framing on cognitive (knowledge test performance), motivational (situational interest), and affective (academic emotions) outcomes in the context of digital history learning. Participants were randomly assigned to one of four conditions in a 2 (Format: video game vs. text) × 2 (Instruction: learning vs. fun) between-subjects design. *Limes,* a video game on the Roman Empire, and its structurally aligned text-version were developed to guarantee consistency in content and engagement mechanisms across conditions (Viccari et al., 2024; Viccari & Bardach, 2025). Based on the literature reviewed above, we formulated a set of hypotheses and more exploratory research questions. All hypotheses and research questions, as well as the methods, were preregistered on aspedicted.org prior to data collection (blinded link for peer-review: https://aspredicted.org/r9f8-f366.pdf).

First, with respect to format effects, we specified the following primary hypotheses: Participants in the more immersive video game condition are expected to report higher situational interest than those in the text condition **(***Hypothesis 1, H1).* In addition, participants in the game condition are expected to score higher on the knowledge test compared to those in the text condition (*Hypothesis 2, H2)*. Participants in the game condition are expected to show more adaptive academic emotions (i.e., higher enjoyment, lower boredom) than those in the text condition (*Hypothesis 3, H3)*.

Second, we examined effects of the instruction. We hypothesized that participants in the learning instruction condition score higher on the knowledge test than those in the fun instruction condition (*Hypothesis 4, H4*).

Third, we investigated the interaction between format and instruction. We expected a significant interaction effect between format (video game vs. text) and instruction (having fun

vs. learning) on knowledge test performance (*Hypothesis 5, H5*). Participants playing the game may require additional instructional support to learn effectively due to its increased complexity and the presence of game elements unrelated to learning (e.g., minigames). Providing explicit instructions to view the game as a learning opportunity could serve as such support by directing their attention to relevant content. In contrast, similar support may be less necessary for learners using the text version, as it contains fewer distractions. Consequently, the effects of a learning instruction are expected to be more pronounced for participants engaging with the game.

In addition, we addressed several (pre-registered) more exploratory research questions without outlining specific hypotheses a priori due to the lack of research on these issues. Specifically, we asked whether there is an effect of the instruction (having fun versus learning) on situational interest (exploratory research question 1), and on academic emotions in terms of engagement and boredom (exploratory research question 2). We also examined whether there is a significant interaction effect between format (game vs text) and instruction (having fun vs learning) on situational interest and on academic emotions (exploratory research questions 3 and 4, respectively).

## Methods

### Participants and design

A total of 143 university students from psychology, teacher education, and sport science programmes took part in the study (107 women, 34 men, 1 non-binary, 1 no response; $M_{age}$= 23.19, $SD$ = 3.95). The study was conducted in a controlled laboratory setting at a German university. Based on an a priori power analysis using G*Power 3.1 (Faul et al., 2009) for an analysis of covariance (fixed effects, main effects, and interactions) in a 2×2 between-subjects design, assuming a medium effect size ($f$ = 0.25), an alpha level of .05, and a power of .80, we aimed to recruit at least $N$ = 128 participants. We collected data from 143 individuals to ensure a sufficiently large sample size and to account for potential exclusions.

Participants were recruited through multiple channels, including flyers posted at the university, email announcements, and the university's platform for study participation, where the study was listed as a standard in-lab study offering course credit for participation as compensation. For a limited period, participants also had the option to receive €20 instead of course credit. To be eligible, participants had to be at least 18 years old, fluent in German, and not enrolled in a history degree program. Individuals with uncorrected severe visual impairments were also excluded.

Participants were randomly assigned to one of four experimental conditions in a 2 (Format: Game vs. text) × 2 (Instruction: Fun vs. learning) between-subjects design. In the *Game–Fun* condition, 37 participants were assigned (27 women, 10 men). The *Game–Learning* condition included 36 participants (29 women, 7 men). The *Text–Fun* condition comprised 35 participants (23 women, 11 men, 1 no response), and the *Text–Learning* condition included 35 participants (28 women, 6 men, 1 non-binary).

**Experimental manipulation and materials**

Participants received condition-specific instructions, depending on their randomly assigned condition (fun vs. learning) prior to engaging with the assigned game or text application (Table 1). The fun- and learning-instruction were matched in content and structure across both formats to ensure comparability. Following the instruction, the assigned application—either the game or the text app—was manually launched by the experiment leader.

**Table 1**

*Overview of the Experimental Instructions by Condition*

| Condition | Instruction |
|---|---|
| Game-Learning | Use the video game *Limes* to learn as much as possible about the content. Explore the game world attentively and focus on the provided information and tasks that can help you learn. |
| Game-Fun | Enjoy the video game *Limes* and have fun! Explore the content and interact with the elements in the way that brings you the most joy. |
| Text-Learning | Use the *Limes* text app to learn as much as possible about the content. Focus on the provided information and tasks that can help you learn. |
| Text-Fun | Enjoy the *Limes* text app and have fun! Explore the content and interact with the elements in the way that brings you the most joy |

Participants engaged with either the educational video game *Limes*, an adventure-style game with strong narrative and entertainment features, or an aligned interactive text-based application. Both formats conveyed identical historical content related to life along the Roman frontier, also known as *Limes*, including themes such as military structures, local settlements, and historical events. The material was divided into five levels, each featuring a main storyline, an optional secondary mission, and supplemental codex entries for historical context (see Figure 1 and Figure 1 for illustrations of the structural alignment across formats).

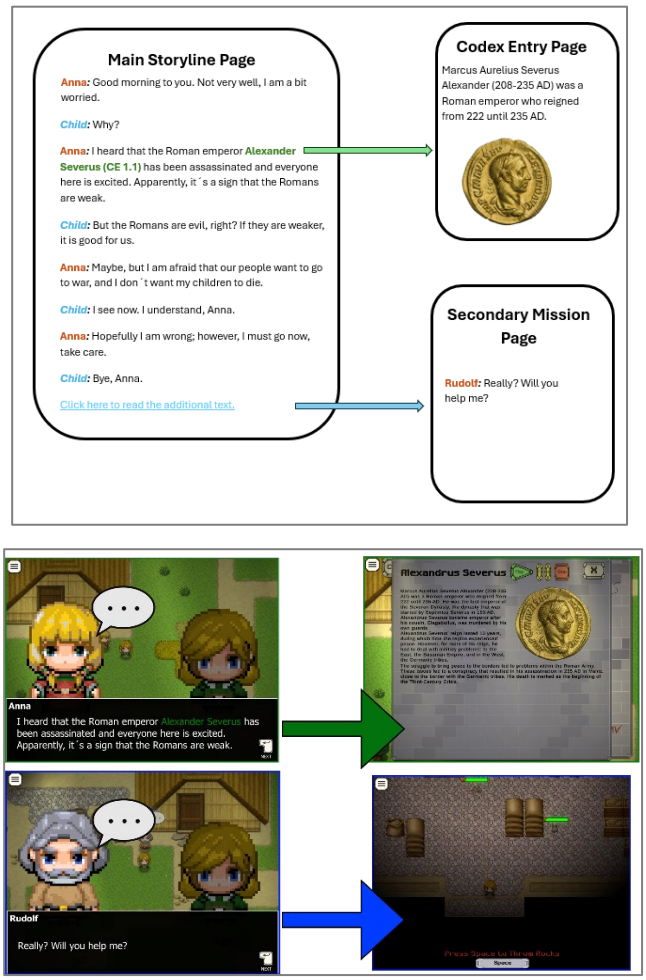
The *Limes* video game (Viccari et al., 2024), developed in Unity (version 2021.3.7f1) as a 2D top-down adventure for Windows and macOS, allowed players to freely navigate virtual

environments, engage in branching dialogues, and interact with optional content. The narrative follows a Germanic child who, after being kidnapped by Roman soldiers, embarks on a complex journey across Roman and Germanic territories. Along the way, the child forges a reluctant alliance—and eventually a friendship—with a Roman centurion named Ariovist. The story unfolds across five levels, covering themes such as Roman military infrastructure, the cultural dynamics between Romans and Germanic tribes, and broader political developments of the third century CE.

Secondary missions offer gameplay elements such as horse racing, stealth sequences, and puzzle-solving, while reinforcing the historical themes presented in the main storyline. Codex entries represent optional historical background information embedded within the game world. These entries—36 in total—were unlocked by interacting with non-playable characters or specific objects and were distributed across the five levels (8 each in Levels 1, 2, and 4; and 6 each in Levels 3 and 5). Each codex entry combined a short, illustrated article and a relevant image (Figure 2), such as a period map or thematic visual. On average, the entries in the German version contained 148 words (range: 59–265), offering concise yet informative insights into topics like Roman emperors, provinces, and historical events. This structure allowed players to explore historically accurate content at their own pace and in alignment with the unfolding story.

**Figure 1**

*Comparison of the interactive text-based application (left) and the video game (right) showing structural alignment across both experimental condition*

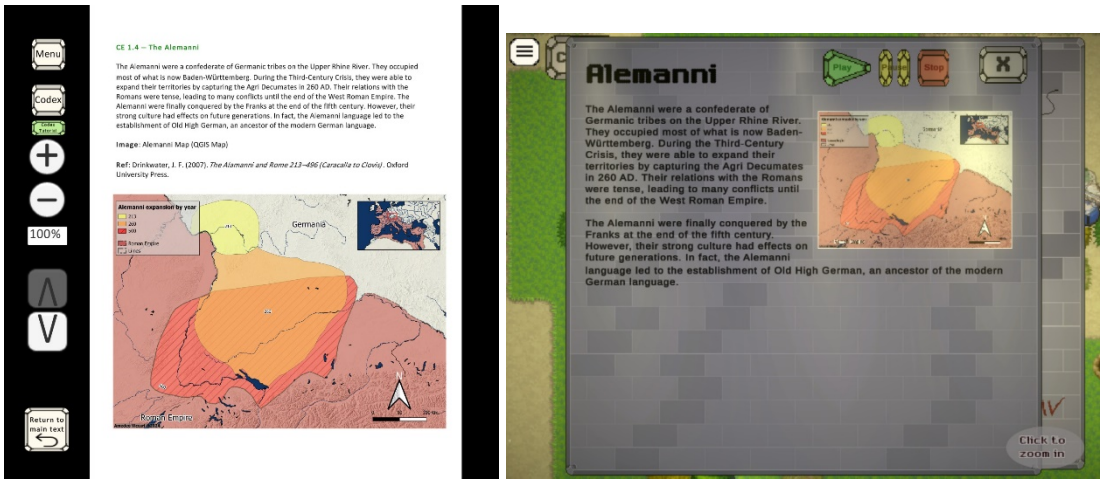



*Note.* The left panel illustrates the general structure of the text application, which presents the Main Storyline Page with clickable links leading to Codex Entry Pages (top) and Secondary Mission Pages (bottom). The right panel shows the same elements of the video game application.

The text-based application (Viccari & Bardach, 2025) was developed in Unity using the same game engine as the video game to ensure structural equivalence. All narrative events, and content locations from the game were preserved within a fixed-sequence interactive format. Secondary missions and Codex entries were integrated as clickable elements embedded within the text, allowing participants to voluntarily explore additional historical information. The text contained 18,061 words, divided into 9,139 words for the main storyline, 2,594 words for secondary missions, and 6,328 words for Codex entries.

**Figure 2**

*Comparison of the same codex entry as shown in the text-based version (left) and in the video game (right)*




In both formats, participants' behaviours were automatically recorded. This included engagement with codex entries, measured by the number of codex clicks and the total time spent viewing codex content; completion of secondary missions; and time-on-task across content sections. Time-on-task was recorded separately for Codex entries and for the main content, which included time spent on the main storyline and secondary missions. On average, participants spent approximately 72.3 minutes interacting with the video game and 63.7 minutes with the text application. In addition, they spent on average 1.51 minutes on the instruction screen before starting the task (see Table 2 for detailed engagement times). Both formats covered identical information about Roman history. Following development, both materials were reviewed by experts in Roman history, educational design, and digital learning to ensure content accuracy, instructional clarity, and usability. We also conducted pilot experiments with university students to ensure that the content was appropriate for the target population.

**Table 2**

*Time spent in Limes (in minutes) by condition*

| | Text_Learning (*n* = 33) | | Text_Fun (*n* = 35) | | Game_Learning (*n* = 36) | | Game_Fun (*n* = 36) | |
|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Instruction Time | 1.91 | 0.57 | 2.21 | 0.68 | 1.65 | 0.50 | 1.64 | 0.53 |
| Codex Entries Time | 18.45 | 9.53 | 18.26 | 10.02 | 9.98 | 9.04 | 11.53 | 9.72 |
| Main Content Time | 45.38 | 9.30 | 45.32 | 9.20 | 62.84 | 12.56 | 60.30 | 10.80 |
| Total Time Codex + Main Content | 63.83 | 16.24 | 63.59 | 14.57 | 72.82 | 13.57 | 71.83 | 13.72 |

## Measures

### Knowledge test

We assessed factual knowledge about Roman history using a self-developed 25-item single-choice test. Each item (e.g., “In which Punic War was Hannibal famously involved?”) presented five response options: four content alternatives (only one correct) and an additional ‘I don’t know’ option. Correct responses contributed one point to the overall sum score (range: 0–25). Because the questions covered separate, factual topics and were coded dichotomously as correct or incorrect, internal consistency was not applicable. In addition, a 5-items subset of the knowledge test were also administered at pretest to assess participants’ prior knowledge (score range: 0–5) to be used as control variable.

### Situational interest

Situational interest was measured using eleven items adapted from Linnenbrink-Garcia et al. (2010), capturing three components: triggered interest (4 items; e.g., “Playing *Limes* is exciting”), maintained interest – feeling (4 items; e.g., “I am enthusiastic about what I learn in *Limes*”), and maintained interest – value (3 items; e.g., “What I learn in *Limes* is valuable to me”). All items were rated on a 7-point Likert scale (1 = does not apply at all, 7 = fully applies). Following Krebs et al. (2024), the two maintained interest components were combined into a single maintained interest scale for analyses. Subscale means were computed for both triggered and maintained interest. Internal consistency was high for triggered interest ($\alpha$ = .86–.92) and

maintained interest ($\alpha$ = .86–.92) across conditions.

**Academic emotions**

Academic emotions were assessed using the short version of the Achievement Emotions Questionnaire (AEQ-S; Bieleke et al., 2021), which includes four items each for enjoyment (e.g., "Engaging with the content of *Limes* brings me joy") and boredom (e.g., "The content of *Limes* is so boring that I catch myself daydreaming"). Participants rated their agreement on a 5-point Likert scale (1 = do not agree at all, 5 = fully agree). Internal consistency was acceptable for enjoyment ($\alpha$ = .67–.86) and boredom ($\alpha$ = .68–.85).

**Pretest background variables and covariates**

In the pretest, we assessed several background variables to be used as potential covariates in the analyses. Prior interest in history was measured using three items (e.g., "How interested are you in Roman history?", adapted from Schumacher et al. 2025) rated on a 7-point Likert scale ($\alpha$ = .66–.79). Prior self-reported historical knowledge ("How much knowledge do you have about the following topics?"), was measured with three items, each rated on a 7-point Likert scale (1 = no knowledge at all, 7 = a great deal of knowledge). The items assessed knowledge of Roman history, ancient history, and history in general (adapted from Schumacher et al., 2025). Responses were averaged to form a composite scale score ($\alpha$ = .68–.85). Participants also reported their average weekly gaming hours.

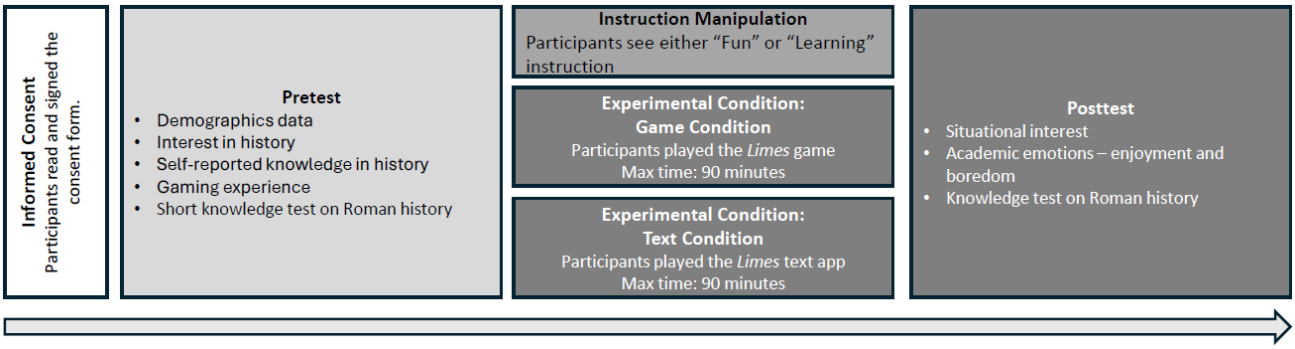
**Procedure**

The experiment was conducted in a laboratory setting at a German university. Upon arrival, participants were welcomed by the experiment leader, informed about the study's procedures and data protection, and provided informed consent. Assignment to the format condition (game vs. text) was randomized prior to participation using a custom Excel macro. Participants then completed the pretest assessments on Unipark, a platform for web-based survey research, via Firefox on university-provided computers. The pretest included demographic and background information, self-reported history knowledge, self-reported interest in history, epistemic curiosity, and a short knowledge test on Roman history. After completing the pretest, participants received the condition-specific instructional framing (fun vs. learning), which was randomized via Unipark's trigger system. After participants read the instructions, the experiment leader manually launched the assigned learning material—either the video game or the text-based application—which were both standalone desktop apps. Participants were instructed to spend a maximum of 90 minutes engaging with the material.

Following the learning phase (or once time had run out), participants returned to the browser to complete the posttest on Unipark, which included measures of situational interest, academic emotions (enjoyment and boredom), and a 25-item factual knowledge test on Roman history. The maximum total duration for each laboratory session was 120 minutes. The study was approved by the University of [blinded for peer-review] ethics board (approval code: [blinded for review]). Figure 3 displays a visualization of the study procedure.

**Figure 3**

*Overview for the studies' experimental procedure*




*Note.* Maximum completion time maximum was 120 min.

## Analyses

We conducted 2 × 2 analyses of covariance (ANCOVAs) to test the main effects of format and instruction, as well as their interaction, on all outcome variables. Correlations between pretest variables and the dependent variables were inspected to determine their inclusion as covariates. All tests of significance were evaluated at the .05 level, and directional hypotheses were assessed using one-tailed tests. We relied on two-tailed tests for all exploratory research questions and exploratory analyses, and for the effects of covariates.

# Results

## Preliminary analyses

As preregistered, participants were to be excluded if their total time spent engaging with the game or text app was below 20 minutes. No participant met this exclusion threshold. However, three participants were excluded due to procedural violations related to the instructional manipulation. Specifically, they failed to stop at the mandatory "stop screen" presented at the end of the pretest. This screen was critical for ensuring that participants read the correct goal-framing instructions (either fun or learning), as it also informed them to notify the experiment leader before proceeding. Without this step, we could not guarantee that the manipulation was delivered and understood as intended. These exclusions ($N = 3$) led to a final

sample of $N = 140$ used for all analyses.

Kruskal–Wallis tests revealed no significant differences in age, $\chi^2(3) = 0.51$, $p = .916$; pretest interest in history, $\chi^2(3) = 1.13$, $p = .770$; self-reported prior knowledge, $\chi^2(3) = 2.42$, $p = .490$; factual historical knowledge, $\chi^2(3) = 1.46$, $p = .691$; or weekly gaming hours, $\chi^2(3) = 2.58$, $p = .462$, across experimental conditions. Chi² tests likewise showed no group differences in gender, $\chi^2(9) = 9.22$, $p = .417$., or gaming experience, $\chi^2(3) = 2.33$, $p = .507$. Table 3 presents descriptive statistics (means, standard deviations) for all outcome variables by condition.

**Table 3**

*Means and standard deviations for outcome variables*

| | Text_Learning ($n$ = 33) | | Text_Fun ($n$ = 35) | | Game_Learning ($n$ = 36) | | Game_Fun ($n$ = 36) | |
|---|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *M* | *SD* | *M* | *SD* |
| Knowledge (Score range: 0–25) | 15.36 | 3.90 | 16.14 | 3.90 | 13.53 | 5.26 | 14.33 | 4.74 |
| Triggered Situational Interest (Scale: 1–7) | 4.48 | 1.38 | 4.38 | 1.43 | 4.42 | 1.22 | 4.89 | 1.12 |
| Maintained Situational Interest (Scale: 1–7) | 4.39 | 1.11 | 4.38 | 1.05 | 4.17 | 1.29 | 4.52 | 1.04 |
| Enjoyment (Scale: 1–5) | 3.20 | 0.75 | 3.10 | 0.87 | 3.36 | 0.87 | 3.47 | 0.65 |
| Boredom (Scale: 1–5) | 1.82 | 0.80 | 1.75 | 0.89 | 1.79 | 0.79 | 1.48 | 0.56 |
| Codex Clicks (Count)[a] | 36.21 | 8.92 | 34.83 | 9.81 | 15.44 | 12.95 | 19.08 | 12.95 |
| Codex Time (Minutes) [a] | 18.45 | 9.53 | 18.26 | 10.02 | 9.98 | 9.04 | 11.53 | 9.72 |

*Note*. [a]Codex clicks and time were used in exploratory analyses.

Following our preregistration, we examined the correlations between pretest variables and the dependent variables to determine whether to include them as covariates. Based on these analyses, we included the following pretest measures as covariates in all ANCOVA models: interest in history, self-reported knowledge, and knowledge test scores (see Appendix A for the full correlation matrix).

**Effects on situational interest**

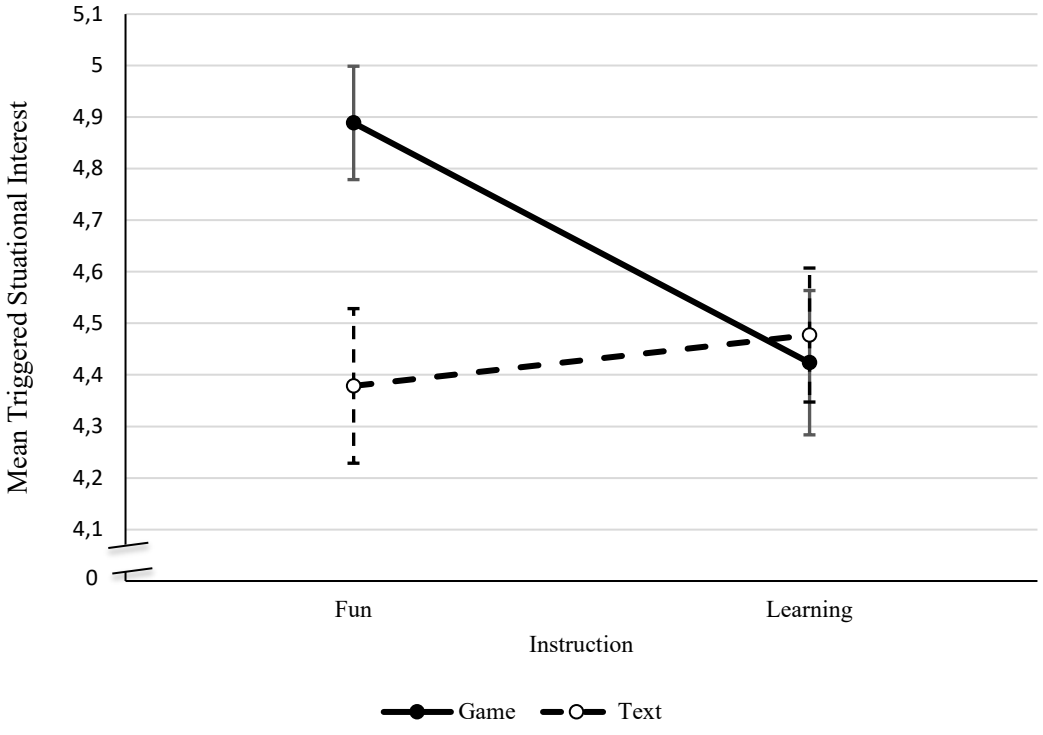
We conducted two separate 2×2 ANCOVAs on triggered and maintained situational interest with format and instruction as between-subjects factors. Pretest interest in history, pretest knowledge test score, and self-reported prior knowledge were included as covariates. The ANCOVA for triggered situational interest revealed a statistically significant main effect of format, $F(1,133) = 2.93$, $p = .045$, $\eta^2_p = .009$, with participants in the game condition reporting

higher levels of triggered interest ($M$ = 4.66, $SD$ = 1.19) than participants in the text condition ($M$ = 4.43, $SD$ = 1.39). There was no significant main effect of instruction (learning: $M$ = 4.45, $SD$ = 1.29; fun: $M$ = 4.64, $SD$ = 1.30), $F$(1,133) = 1.47, $p$ = .227, $\eta^2_p$ = .006. The format × instruction interaction was also not significant, $F$(1,133) = 1.24, $p$ = .267, $\eta^2_p$ = .009. Among the covariates, pretest interest in history significantly predicted triggered interest, $F$(1,133) = 7.39, $p$ = .007, $\eta^2_p$ = .07.

For maintained situational interest, there were no significant main effects of format (game: $M$ = 4.35, $SD$ = 1.18; text: $M$ = 4.39, $SD$ = 1.07), $F$(1,133) = 0.33, $p$ = .284, $\eta^2_p$ < .01, or instruction (learning $M$ = 4.28, $SD$ = 1.21; fun: $M$ = 4.45, $SD$ = 1.04), $F$(1,133) = 0.73, $p$ = .395, $\eta^2_p$ = .008. The format × instruction interaction was also not significant, $F$ 1,133) = 0.50, $p$ = .480, $\eta^2_p$ = .004. Again, pretest interest in history was a significant covariate, $F$(1,133) = 22.26, $p$ < .001, $\eta^2_p$ = .19.

**Figure 4**

*Situational interest (means) by format (game vs. text) and instruction (fun vs. learning)*




*Note.* Situational interest was measured with 11 items adapted from Linnenbrink-Garcia et al. (2010). Triggered interest (4 items) and maintained interest (7 items) were rated on a 7-point Likert scale (1 = does not apply at all, 7 = fully applies). Error bars represent standard errors of the mean.

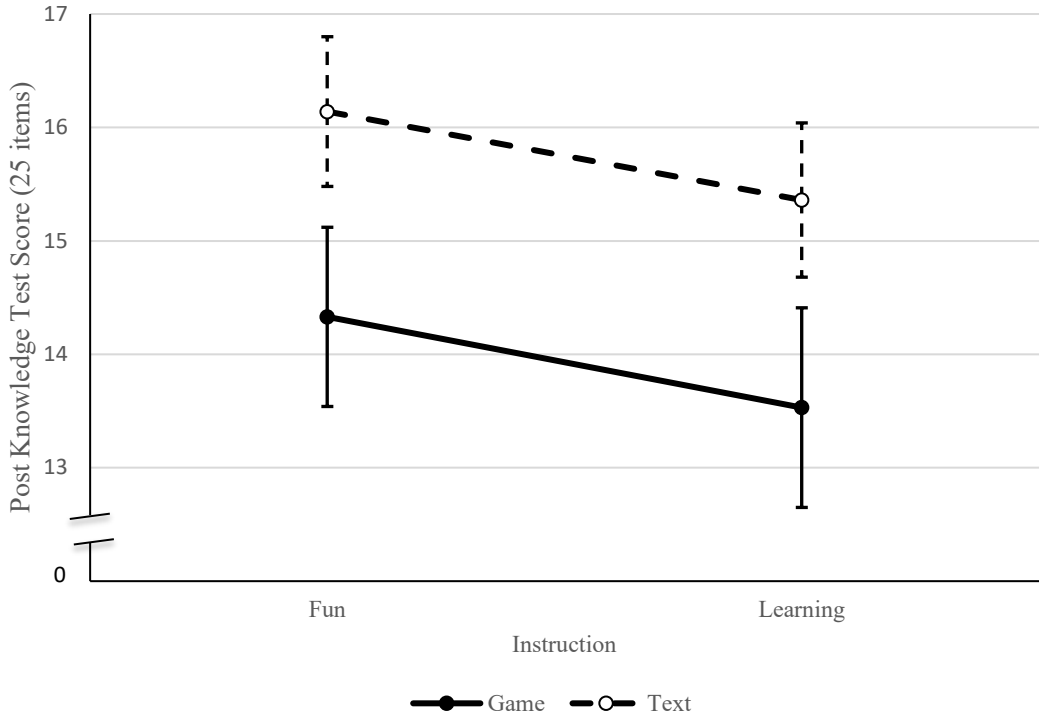
**Effects on knowledge test performance**

We conducted a 2×2 ANCOVA on knowledge test performance with format (game vs. text) and instruction (fun vs. learning) as between-subjects factors. Pretest interest in history, the pretest knowledge test score and self-reported prior knowledge were included as covariates. Although the main effect of format (game: $M$ = 13.9, $SD$ = 4.99; text: $M$ = 15.8, $SD$ = 3.89) reached significance in a two-tailed test, $F(1, 133) = 4.64$, $p = .033$, $\eta^2_p = .041$, and would have reached significance in a one-tailed test had it been in the predicted direction, it was not significant under the preregistered one-tailed test, which assumed higher knowledge test scores in the game condition than in the text condition. For interested readers and to inform future research, the means are reported in Figure 5 for completeness.

There was no statistically significant main effect of instruction (learning: $M$ = 14.4, $SD$ = 4.72; fun: $M$ = 15.2, $SD$ = 4.41), $F(1,133) = 0.13$, $p = .722$, $\eta^2_p = .01$ (H4), nor a significant format x instruction interaction, $F(1,133) = 0.07$, $p = .799$, $\eta^2_p < .01$ (H5). Among the covariates, only pretest interest in history statistically significantly predicted knowledge test performance, $F(1,133) = 6.25$, $p = .014$, $\eta^2_p = .073$.

**Figure 5**

*Knowledge test performance (means) by format (game vs. text) and instruction (fun vs. learning)*




*Note.* Posttest scores are based on the factual knowledge test (25 items). Full range of the test score was 0 (no correct answers) to 25 (all answers correct). Error bars represent standard errors

of the mean.

**Effects on academic emotions**

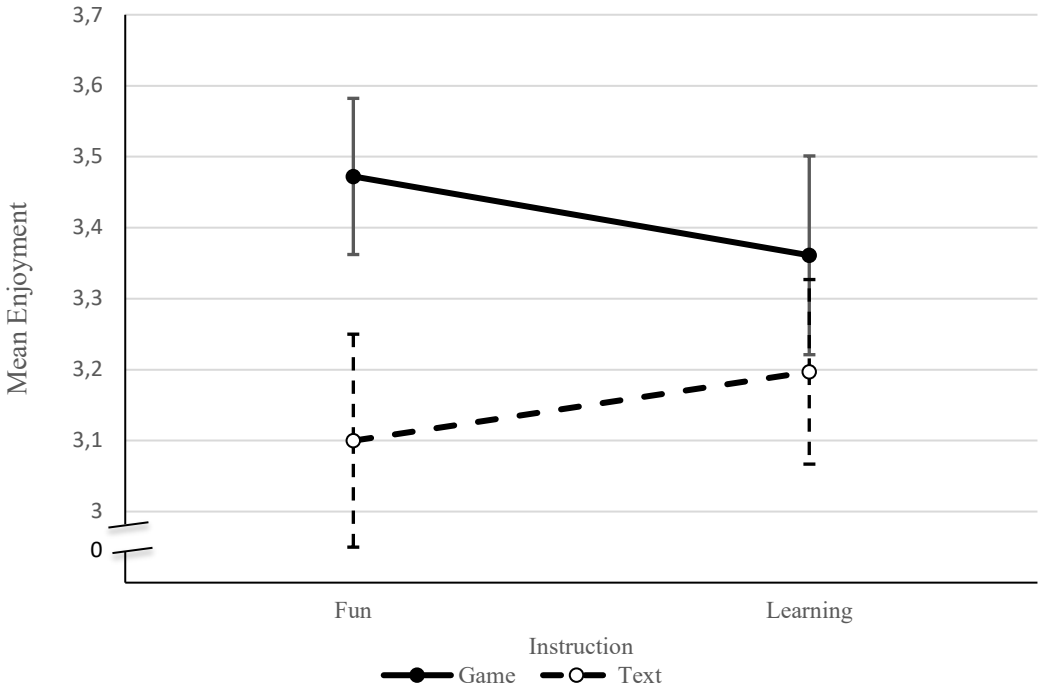
We conducted two 2×2 ANCOVAs on enjoyment and boredom with format and instruction as between-subjects factors. Covariates included were pretest interest in history, the pretest knowledge test score, self-reported prior knowledge, and weekly gaming hours.

For enjoyment, the ANCOVA revealed a statistically significant main effect of format, $F(1,133) = 4.23$, $p = .021$, $\eta^2_p = .029$, with participants in the game condition reporting more enjoyment ($M = 3.42$, $SD = 0.76$) than those in the text condition ($M = 3.15$, $SD = 0.81$), in line with H3 (see Figure 6). There was no main effect of instruction (learning: $M = 3.28$, $SD = 0.81$; fun: $M = 3.29$, $SD = 0.78$), $F(1,133) = 0.03$, $p = .870$, $\eta^2_p < .001$, and no format × instruction interaction, $F(1,133) = 0.32$, $p = .572$, $\eta^2_p = .002$. The covariate pretest interest in history significantly predicted enjoyment, $F(1,133) = 9.95$, $p = .002$, $\eta^2_p = .058$.

For boredom, the ANCOVA revealed no statistically significant main effects of format (game: $M = 1.64$, $SD = 0.70$; text: $M = 1.78$, $SD = 0.84$), $F(1,133) = 2.18$, $p = .071$, $\eta^2_p = .009$, or instruction (learning: $M = 1.80$, $SD = 0.79$; fun: $M = 1.61$, $SD = 0.75$), $F(1,133) = 2.22$, $p = .139$, $\eta^2_p = .016$. The format × instruction interaction was also not significant, $F(1,133) = 0.64$, $p = .425$, $\eta^2_p = .005$. None of the covariates were significant predictors of boredom.

**Figure 6**

*Enjoyment ratings (means) by format (game vs. text) and instruction (fun vs. learning)*




*Note*. Enjoyment scores are based on the AEQ-S Enjoyment subscale (4 items). Full scale range was 1 (do not agree at all) to 5 (fully agree). Error bars represent standard errors of the mean.

**Exploratory analyses**

We conducted exploratory analyses using behavioural data (interactions with the codex entries) to explore participants' interactions with the codex and their relation to posttest knowledge performance, and, potentially, to shed light on reasons behind the higher knowledge test scores in the text condition. We conducted ANCOVAs, correlation analyses, and multiple regression analyses while controlling for prior knowledge and relevant background variables. Codex clicks refer to the number of times participants opened individual entries within the codex interface during the task, while codex time reflects the total time (in seconds) spent interacting with codex content. Although these analyses were not preregistered, we included them to investigate how engagement with supplemental in-game information (i.e., the codex) might contribute to learning outcomes, as the codex was designed to provide historically accurate content in both formats.

***Codex usage by format.*** We conducted two separate 2×2 ANCOVAs with codex clicks and codex time as dependent variables. The between-subjects factors were format (game vs. text) and instruction (fun vs. learning). Based on prior correlation analyses, we included self-reported prior knowledge as a covariate, given its significant association with posttest

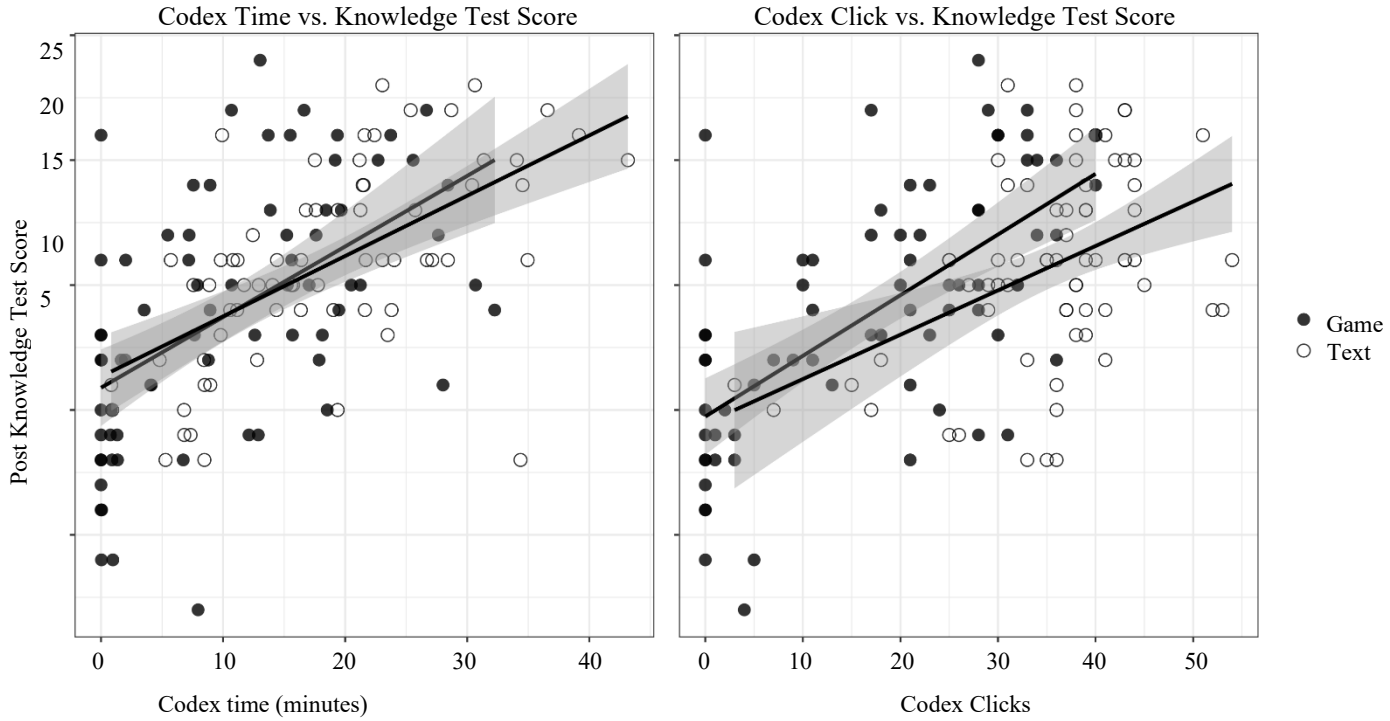
knowledge scores (see Appendix A for full correlation matrix). This approach allowed us to control for individual differences in knowledge when analysing codex engagement across experimental conditions.

For codex clicks, the ANCOVA revealed a significant main effect of format, $F(1,135) = 34.12$, $p < .001$, $\eta^2_p = .395$, indicating that participants in the text condition opened significantly more codex entries ($M = 35.5$, $SD = 9.34$) than those in the game condition ($M = 17.3$, $SD = 13.0$). Similarly, for codex time, the ANCOVA revealed a significant main effect of format, $F(1,135) = 8.68$, $p = .004$, $\eta^2_p = .139$. Participants in the text condition spent more time engaging with the codex ($M = 18.36$ minutes, $SD = 9.81$) than those in the game condition ($M = 10.25$ minutes, $SD = 9.01$). Again, there were no statistically significant effects of instruction on codex clicks (learning: $M = 25.4$, $SD = 15.3$; fun: $M = 26.8$, $SD = 13.9$), $F(1,135) = 1.67$, $p = .198$, $\eta^2_p = .003$, or on codex time (learning: $M = 14.0$ min, $SD = 10.2$; fun: $M = 14.8$ min, $SD = 10.4$), $F(1,135) = 0.49$, $p = .484$, $\eta^2_p = .001$.

***Relation to posttest knowledge.*** Correlation analyses revealed significant positive associations between posttest knowledge test scores and both codex clicks ($r = .554$, $p < .001$) and codex time ($r = .580$, $p < .001$). To examine these relationships while controlling for self-reported prior knowledge, we conducted two multiple regression analyses. In the first model, codex clicks significantly predicted posttest knowledge scores ($b = 0.18$, $p < .001$). In the second model, codex time also significantly predicted posttest knowledge scores ($b = 0.26$, $p < .001$). These results indicate that participants who interacted more frequently and for longer durations with the Codex achieved higher posttest knowledge scores, above and beyond the effects of their self-reported prior knowledge.

**Figure 7**

*Associations between codex interaction and posttest knowledge performance*




*Note.* The left panel shows the relationship between total codex time (in minutes) and posttest knowledge test scores; the right panel shows the relationship between the number of codex clicks and posttest knowledge test scores. Posttest knowledge was assessed with a factual knowledge test (25 items; range: 0 = no correct answers to 25 = all correct answers). Shaded areas represent 95% confidence intervals.

## Discussion

The goal of this study was to explore whether differences in the format and instructional framing of history learning materials affect academic outcomes. Specifically, we examined how instructional framing (learning vs. fun) and format (video game vs. text) influence knowledge acquisition, situational interest, and academic emotions in the context of digital history learning. Using a 2×2 between-subjects design, we compared a historically accurate video game on the Roman Empire with a structurally aligned interactive text application, both presenting identical historical content.

**Format effects**

Situational interest plays a central role in motivating learners to engage with educational content, especially in game-based learning environments. In line with this, prior studies have shown that games can elicit higher situational interest than text formats (e.g., Wouters et al., 2013). Our results revealed a significant difference between the two conditions regarding triggered interest, but not with regard to maintained interest, thus partially confirming Hypothesis 1. This finding indicates that game-based formats may be effective in capturing learners' initial attention and curiosity (triggered interest), albeit it is important to note that the observed effect on triggered interest was small and only just reached statistical significance. A possible reason for the lack of significant differences with respect to maintained interest may be the structural alignment between the two formats: the interactive text application offered similar features—such as clickable content, codex entries, and optional exploration paths—which may have effectively preserved learners' sense of autonomy and engagement, thereby sustaining learners' interest similarly to the game format. This finding aligns with Gui et al. (2023), who demonstrated that structural equivalence, especially in terms of interactivity and navigation, can reduce differences between media types.

Contrary to our expectations (Hypothesis 2), participants in the video game condition did not perform significantly better than those in the text condition. Instead, the knowledge test scores of the text condition were higher than those of the video game condition. This challenges the belief that digital games inherently enhance factual learning, and aligns with previous findings showing that games do not consistently outperform traditional formats on knowledge retention tasks (Adams et al., 2012; Brom et al., 2019). One potential explanation lies in too little active goal-directed processing in the game condition (Eitel et al., 2025): the immersive and exploratory nature of the game did not particularly support or reward students to pursue the goal of being best prepared for the knowledge test. Navigating the in-game environment was rewarding in itself (also shown by higher triggered situational interest in game condition), but it did not specifically reward for processing historical content in the codex entries that was required for the knowledge test. Active goal-directed processing, for the goal of being best prepared for the knowledge test, was thus lower in the game than in the text condition. Behavioural tracking supports this interpretation. Although game participants spent more time on the task overall (≈72 min) compared to text participants (≈64 min), they devoted substantially less time to codex entries (≈10–11 min vs. ≈18 min) (see Table 2). As our exploratory analyses revealed that codex engagement was strongly associated with knowledge

performance, this suggests that additional time in the game was likely devoted to navigation and interactive mechanics rather than to processing historical information. The text condition's linearity and interface design seems to have cognitively foregrounded codex content, making it more visible and accessible during the learning process. In theoretical terms, the linearity increased focus support for students to increase active goal-directed processing, in turn leading to stronger learning gains (Eitel et al., 2025). In contrast, the game condition—despite offering identical codex material—saw lower engagement with codex entries, possibly because its visual and interactive richness drew attention toward gameplay tasks and away from optional text-based resources (lower focus support; Eitel et al., 2025). These differences suggest that in content-rich environments, ease of access and integration of supplemental materials are critical for learning outcomes (see also Habgood & Ainsworth, 2011). Moreover, the game's non-linear structure may have led some learners to skip or overlook key information that was more directly presented in the linear text format.

The findings revealed statistically significant format effects on enjoyment, with participants in the game condition reporting higher enjoyment than those in the text condition, thus supporting our hypothesis on effects on enjoyment (see Hypothesis 3). This is consistent with previous findings that associate game-based environments with greater emotional engagement and immersion (Sabourin & Lester, 2014; Taub et al., 2020). In addition, given that the text condition was structurally aligned with the game and therefore likely more engaging than the static text formats typically employed in format-comparison studies, the higher enjoyment ratings observed in the game condition are of particular note. However, boredom levels did not significantly differ between the game and text conditions, which contradicts our respective assumptions on boredom as outlined in Hypothesis 3 (i.e., significantly lower levels of boredom in the game condition). This pattern for boredom may reflect the overall quality and coherence of both interventions, which offered interactive elements and rich historical content. While the game format seemed to increase enjoyment, boredom levels were low across both groups. This may suggest that well-designed and engaging content can counteract boredom, no matter the format.

**Effects of instructional framing**

Contrary to our Hypothesis 4 and prior findings (e.g., Erhel & Jamet, 2019; Hu et al., 2025), we did not obtain any statistically significant effects of instructional framing (learning vs. fun) on knowledge test performance. Also, no significant effects for interest or academic emotions emerged (exploratory research questions 1 and 2, respectively).

One possible explanation is that the instructional framing, although implemented in line with prior studies on its effects (Erhel & Jamet, 2013, 2019), may have been too subtle or not sufficiently reinforced during the task. Although participants received instructions tailored to their condition, the lack of in-task reminders or goal-specific scaffolds may have made the instructional guidance less impactful (Pilegard & Mayer, 2016). Given that participants spent on average more than a minute on the instruction screen (with longer times in the text than in the game condition, see Table 2), it seems unlikely that null effect was attributable to insufficient reading time.

It is also possible that learners in our university sample approached both frames with a similar learning goal orientation. Although we did not directly assess learning goals (e.g., (Elliot & McGregor, 2001; Senko, 2016) pretest interest in history was moderate to high across conditions, suggesting that participants were already inclined to engage with the material. Moreover, the relationship between motivationally supportive environments and students' adoption of learning goals is well established (for a meta-analysis, see Bardach et al., 2020). Hence, engaging with Roman history in the motivationally supportive digital environment of our study content may have further supported participants' learning goals, irrespective of the instructional framing. Overall, these findings highlight the challenge of implementing brief, one-time instructional manipulations in controlled lab settings—particularly with content-rich and exploratory materials—pointing to the need for more sustained or embedded framing strategies in future research.

**Interaction effects between format and instruction**

No significant interaction between content format and instructional framing was found for knowledge test performance. This result contradicts our Hypothesis 5, which assumed that the game format—because of its higher cognitive demands and potential for distraction—was expected to benefit more from an explicit learning instruction that could help focus attention and reduce extraneous processing. Similar hypotheses have been proposed in the literature, where it is often suggested that open-ended or game-like environments may require stronger instructional scaffolding to guide attention and processing (e.g., Clark et al., 2016; Erhel & Jamet, 2013). Furthermore, we did not find significant interaction effects for any of the other outcomes (see exploratory research questions 2 and 3, concerning interest and emotions, respectively). With respect to triggered situational interest and enjoyment, this indicates that the significant effects of format are consistent across the different types of instruction.

One possible explanation is that the high structural alignment between the two formats—each offering similar access to features such as optional missions and embedded learning cues—reduced the likelihood of diverging outcomes, as participants may have engaged with the material in similarly self-directed ways, irrespective of the instructional framing. Additionally, it is possible that the 90-minute task duration allowed participants enough time to develop their own learning strategies, which may have reduced the impact of the initial instructional framing and its interaction with format (McCrudden et al., 2010). As discussed in the section above, the subtlety of the framing manipulation may also help explain the absence of interaction effects. Taken together, the absence of interaction effects supports the idea that format and instruction may operate independently—or that their interplay may depend on more nuanced design features. For instance, embedding goal-relevant actions directly into gameplay could enhance the effects of instructional framing and lead to different instructional effects depending on the format.

**The role of engaging with codex entries**

Finally, the results from our exploratory analyses focusing on codex entries showed that game participants engaged less with codex entries than text participants. Importantly, greater codex use was strongly linked to higher knowledge scores. This suggests that codex engagement, when it occurs, meaningfully relates to learning outcomes. In the text condition, where codex entries were accessed more frequently (≈35 clicks; ≈18 minutes), the visibility and accessibility of supplementary content appeared to sustain such engagement with the codex. From a theoretical standpoint, these findings seem aligned with the Focus-Support-Reward-Squares model (Eitel et al., 2025), postulating that instruction needs to provide at least a medium degree of focus support to stimulate effective learning. Historically rich, exploratory environments, are challenging in this regard, because their interactive mechanics may divert attention from core content unless supportive resources are tightly integrated into the task flow. In the best case, interactive mechanics feel rewarding for students, are useful to proceed in the game and to internalize core learning contents. This resonates with the concept of intrinsic integration (Habgood & Ainsworth, 2011), highlighting that learning elements are most effective when they are functionally necessary for success within the game environment. This makes active goal-directed processing, and thus good learning outcomes, most likely (Eitel et al., 2025). The lower codex engagement in the game condition—despite identical accessibility across formats—suggests that optional resources can be overlooked when not structurally embedded in progression. This finding is also in accordance with the goal-focusing model

(McCrudden & Schraw, 2007), which posits that sustained, context-embedded cues are critical. Furthermore, the results resonate with

In practical terms, these insights suggest that designers of educational games and interactive texts should integrate learning resources into the core task flow rather than positioning them as optional add-ons. Embedding knowledge elements into central mechanics, coupling them with in-task prompts, and making them consequential for progress can ensure that learning goals remain salient without disrupting immersion. Similarly, text-based formats can benefit from maintaining high visibility and accessibility of supplementary materials, for example, through persistent side panels or embedded pop-ups that keep resources cognitively foregrounded. In *Limes* specifically, future versions could integrate codex content more directly into the gameplay loop—for example, by requiring players to consult codex entries to solve puzzles, decode clues, or unlock gates during missions. Such integration could also take the form of challenges where survival or progression depends on historical knowledge, such as escaping danger or defeating an opponent by applying information from the codex. However, careful planning is essential to prioritise the historical information most critical to the educational goals. Overloading players with too many codex-dependent tasks risks diluting the game's purpose and reducing enjoyment and motivation.

**Limitations and directions for future research**

While this study offers valuable insights into digital history learning, several limitations and, related to that, future research directions, should be acknowledged. First, the instructional framing was likely not persistent enough throughout the intervention to influence academic outcomes. Prior studies suggest that goal instructions are more effective when repeated or embedded within the task environment (McCrudden et al., 2010; Pilegard & Mayer, 2016). Future designs should consider integrating adaptive scaffolding, in-task prompts, or ongoing goal cues to increase the impact of instructional framing. Second, the study assessed all outcomes at the end of the session in the lab. While this approach captures immediate effects, it precludes gaining insights into longer-term retention or motivational impact. Follow-up assessments in future studies could determine whether observed effects persist, or, in the case of non-significant post-test effects, whether effects may instead become detectable at the follow-up assessment. It would also be important to assess whether learners transfer historical knowledge to new contexts—an important benchmark in history education (Gestsdóttir et al., 2018; van Boxtel & van Drie, 2018). Third, although our findings are promising, replications of our study are needed to draw more confident conclusions, particularly given that some of the

observed effects were relatively small. Fourth, the sample consisted exclusively of university students. While appropriate for the context of our study, these participants likely possess more advanced self-regulation skills and may approach learning tasks with greater autonomy, factors that could affect how they engaged with the task. Thus, we envision future research to test *Limes* with learners from different age groups (e.g., older adults, school-aged learners, Benavides-Varela et al., 2020; Yılmaz et al., 2022) whose cognitive, motivational, and attentional characteristics may differ markedly from university students. Finally, while log data revealed a strong association between codex engagement and knowledge test performance, these behavioural traces alone cannot fully explain why some participants engaged more than others. To complement logfile analysis, future work should incorporate qualitative feedback (e.g., interviews or short surveys) to capture learners' perceptions of optional tools. This would provide a richer understanding of how design features shape codex use, particularly among younger users who may interpret relevance and usability differently.

**Conclusions**

The present study examined differences between format and instructional framing using a historically accurate video game (*Limes*) on the Roman Empire and a structurally aligned interactive text application. The findings highlight the potential of video games to foster positive academic emotions and spark learners' interest, while also pointing to potential cognitive limitations of video games. At the same time, our results underscore the importance of meaningfully integrating pedagogical content into the task environment. This may be particularly critical for history education, where balancing narrative immersion with content learning presents unique pedagogical challenges (Pilegard & Mayer, 2016). For games, this may involve embedding knowledge resources (such as codex entries) into core mechanics; for texts, it requires maintaining high visibility and accessibility of such resources. Overall, this study contributes to the growing body of experimental research on digital learning games and digital history learning by offering a controlled comparison of identical content delivered through game and text formats, while also shedding light on the role of instructional framing.

## Appendix A

## Table A1

*Overview of correlations between study variables*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Interest in History (Pretest) | 1.00 | | | | | | | | | | | |
| 2. Self-report Knowledge (Pretest) | 0.63*** | 1.00 | | | | | | | | | | |
| 3. Knowledge Test (Pretest) | 0.50*** | 0.55*** | 1.00 | | | | | | | | | |
| 4. Weekly Gaming Hours (Pretest) | 0.03 | -0.01 | 0.00 | 1.00 | | | | | | | | |
| 5. Interest in History (Posttest) | 0.81*** | 0.50*** | 0.38*** | -0.03 | 1.00 | | | | | | | |
| 6. Knowledge Test (Posttest, 25 items) | 0.44*** | 0.38*** | 0.35*** | 0.01 | 0.41*** | 1.00 | | | | | | |
| 7. Triggered Situational Interest | 0.32*** | 0.19 | 0.19 | -0.05 | 0.48*** | 0.09 | 1.00 | | | | | |
| 8. Maintained Situational Interest | 0.55*** | 0.36*** | 0.28** | -0.06 | 0.74*** | 0.33*** | 0.73*** | 1.00 | | | | |
| 9. Enjoyment | 0.42*** | 0.32*** | 0.29** | -0.07 | 0.57*** | 0.17 | 0.85*** | 0.75*** | 1.00 | | | |
| 10. Boredom | -0.23* | -0.15 | -0.14 | 0.15 | -0.37*** | -0.25** | -0.59*** | -0.45*** | -0.58*** | 1.00 | | |
| 11. Codex Clicks | 0.12 | 0.02 | -0.03 | 0.08 | 0.18 | 0.56*** | -0.10 | 0.06 | -0.11 | 0.05 | 1.00 | |
| 12. Codex Time (sec) | 0.10 | -0.02 | -0.04 | 0.06 | 0.16 | 0.61*** | -0.06 | 0.11 | -0.07 | -0.03 | 0.76*** | 1.00 |

*Note.* *Correlation is significant at the .05 level (two-tailed); **Correlation is significant at the .01 level (two-tailed); ***Correlation is significant at the .001 level (two-tailed).

## References

Adams, D. M., Mayer, R. E., MacNamara, A., Koenig, A., & Wainess, R. (2012). Narrative games for learning: Testing the discovery and narrative hypotheses. *Journal of Educational Psychology*, *104*(1), 235–249. https://doi.org/10.1037/a0025595

Akkerman, S., Admiraal, W., & Huizenga, J. (2009). Storification in History education: A mobile game in and about medieval Amsterdam. *Computers & Education*, *52*, 449–459. https://doi.org/10.1016/j.compedu.2008.09.014

All, A., Castellar, E. P. N., & Looy, J. V. (2014). Measuring Effectiveness in Digital Game-Based Learning: A Methodological Review. *International Journal of Serious Games*, *1*(2), Article 2. https://doi.org/10.17083/ijsg.v1i2.18

Bardach, L., & Murayama, K. (2025). The role of rewards in motivation—Beyond dichotomies. *Learning and Instruction*, *96*, 102056. https://doi.org/10.1016/j.learninstruc.2024.102056

Bardach, L., Oczlon, S., Pietschnig, J., & Lüftenegger, M. (2020). Has achievement goal theory been right? A meta-analysis of the relation between goal structures and personal achievement goals. *Journal of Educational Psychology*, *112*(6), 1197–1220. https://doi.org/10.1037/edu0000419

Barton, K. C., & Levstik, L. S. (2004). *Teaching history for the common good*. Lawrence Erlbaum Associates Publishers.

Benavides-Varela, S., Zandonella Callegher, C., Fagiolini, B., Leo, I., Altoè, G., & Lucangeli, D. (2020). Effectiveness of digital-based interventions for children with mathematical learning difficulties: A meta-analysis. *Computers & Education*, *157*, 103953. https://doi.org/10.1016/j.compedu.2020.103953

Bieleke, M., Gogol, K., Goetz, T., Daniels, L., & Pekrun, R. (2021). The AEQ-S: A short version of the Achievement Emotions Questionnaire. *Contemporary Educational Psychology*, *65*, 101940. https://doi.org/10.1016/j.cedpsych.2020.101940

Brom, C., Stárková, T., Bromová, E., & Děchtěrenko, F. (2019). Gamifying a Simulation: Do

a Game Goal, Choice, Points, and Praise Enhance Learning? *Journal of Educational Computing Research*, *57*(6), 1575–1613. https://doi.org/10.1177/0735633118797330

Chapman, A. (2016). *Digital Games as History: Adam Chapman*. https://www.book2look.com/book/DdgRKX4PQ7

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital Games, Design, and Learning: A Systematic Review and Meta-Analysis. *Review of Educational Research*, *86*(1), 79–122. https://doi.org/10.3102/0034654315582065

Clyde, J., Hopkins, H., & Wilkinson, G. (2012). Beyond the "Historical" Simulation: Using Theories of History to Inform Scholarly Game Design. *Loading...*, *6*(9), Article 9. https://journals.sfu.ca/loading/index.php/loading/article/view/105

Crymble, A. (2021). *Technology and the Historian: Transformations in the Digital Age*. University of Illinois Press. https://doi.org/10.5406/j.ctv1k03s73

de Ribaupierre, S., Kapralos, B., Haji, F., Stroulia, E., Dubrowski, A., & Eagleson, R. (2014). Healthcare Training Enhancement Through Virtual Reality and Serious Games. In M. Ma, L. C. Jain, & P. Anderson (Eds.), *Virtual, Augmented Reality and Serious Games for Healthcare 1* (pp. 9–27). Springer. https://doi.org/10.1007/978-3-642-54816-1_2

Eitel, A., Krebs, M.-C., & Schöne, C. (2025). Introducing the FoRe-Squares Model: Focus Support and Reward as Key Predictors of Effective Technology-Augmented Instruction. *Educational Psychology Review*, *37*(3), 67. https://doi.org/10.1007/s10648-025-10045-z

Elliot, A. J., & McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology*, *80*(3), 501–519. https://doi.org/10.1037/0022-3514.80.3.501

El-shara, I. (2015). *Learning and Teaching between Enjoyment and Boredom as Realized by the Students: A Survey from the Educational Field*. ResearchGate.

Erhel, S., & Jamet, E. (2013). Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & Education*, *67*, 156–167.

https://doi.org/10.1016/j.compedu.2013.02.019

Erhel, S., & Jamet, E. (2019). Improving instructions in educational computer games: Exploring the relations between goal specificity, flow experience and learning outcomes. *Computers in Human Behavior*, *91*, 106–114. https://doi.org/10.1016/j.chb.2018.09.020

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Firaxis Games. (1991). *Civilization (series)* [Computer software].

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Comput. Entertain.*, *1*(1), 20. https://doi.org/10.1145/950566.950595

Gestsdóttir, S. M., van Boxtel, C., & van Drie, J. (2018). Teaching historical thinking and reasoning: Construction of an observation instrument. *British Educational Research Journal*, *44*(6), 960–981. https://doi.org/10.1002/berj.3471

Gui, Y., Cai, Z., Yang, Y., Kong, L., Fan, X., & Tai, R. H. (2023). Effectiveness of digital educational game and game design in STEM learning: A meta-analytic review. *International Journal of STEM Education*, *10*(1), 36. https://doi.org/10.1186/s40594-023-00424-9

Habgood, M. P. J., & Ainsworth, S. E. (2011). Motivating Children to Learn Effectively: Exploring the Value of Intrinsic Integration in Educational Games. *Journal of the Learning Sciences*, *20*(2), 169–206. https://doi.org/10.1080/10508406.2010.508029

Hidi, S., & Renninger, K. A. (2006). The Four-Phase Model of Interest Development. *Educational Psychologist*, *41*(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

Hu, Y., Wouters, P., van der Schaaf, M., & Kester, L. (2025). Timing of information presentation matters: Effects on secondary school students' cognition, motivation and emotion in game-based learning. *British Journal of Educational Technology*, *56*(1),

318–338. https://doi.org/10.1111/bjet.13510

Krebs, M.-C., Braschoß, K., & Eitel, A. (2024). Does watching an explainer video help learning with subsequent text? – Only when prompt-questions are provided. *Learning and Instruction*, *94*, 101988. https://doi.org/10.1016/j.learninstruc.2024.101988

Kusuma, G. P., Putera Suryapranata, L. K., Wigati, E. K., & Utomo, Y. (2021). Enhancing Historical Learning Using Role-Playing Game on Mobile Platform. *Procedia Computer Science*, *179*, 886–893. https://doi.org/10.1016/j.procs.2021.01.078

Linnenbrink-Garcia, L., Durik, A. M., Conley, A. M., Barron, K. E., Tauer, J. M., Karabenick, S. A., & Harackiewicz, J. M. (2010). Measuring Situational Interest in Academic Domains. *Educational and Psychological Measurement*, *70*(4), 647–671. https://doi.org/10.1177/0013164409355699

Mayer, R. E. (2014). Cognitive Theory of Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 43–71). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.005

McCall, J. (2016). Teaching History With Digital Historical Games: An Introduction to the Field and Best Practices. *Simulation & Gaming*, *47*(4), 517–542. https://doi.org/10.1177/1046878116646693

McCall, J. (2023). *Gaming the Past: Using Video Games to Teach Secondary History* (2nd ed.). Routledge.

McCrudden, M. T., Magliano, J. P., & Schraw, G. (2010). Exploring how relevance instructions affect personal reading intentions, reading goals and text processing: A mixed methods study. *Contemporary Educational Psychology*, *35*(4), 229–241. https://doi.org/10.1016/j.cedpsych.2009.12.001

McCrudden, M. T., & Schraw, G. (2007). Relevance and Goal-Focusing in Text Processing. *Educational Psychology Review*, *19*(2), 113–139. https://doi.org/10.1007/s10648-006-9010-7

Murayama, K., FitzGibbon, L., & Sakaki, M. (2019). Process Account of Curiosity and Interest:

A Reward-Learning Perspective. *Educational Psychology Review*, *31*(4), 875–895. https://doi.org/10.1007/s10648-019-09499-9

Ortiz, S. (2014). Video game self-efficacy and its effect on training performance. *Electronic Theses and Dissertations*. https://stars.library.ucf.edu/etd/4710

Paintbucket Games. (2025). *The Darkest Files* [Computer software]. https://store.steampowered.com/app/2058730/The_Darkest_Files/

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

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

Pilegard, C., & Mayer, R. E. (2016). Improving academic learning from computer-based narrative games. *Contemporary Educational Psychology*, *44–45*, 12–20. https://doi.org/10.1016/j.cedpsych.2015.12.002

Ricci, K. E., Salas, E., & Cannon-Bowers, J. A. (1996). Do Computer-Based Games Facilitate Knowledge Acquisition and Retention? *Military Psychology*, *8*(4), 295–307. https://doi.org/10.1207/s15327876mp0804_3

Riopel, M., Nenciovici ,Lucian, Potvin ,Patrice, Chastenay ,Pierre, Charland ,Patrick, Sarrasin ,Jérémie Blanchette, & and Masson, S. (2019). Impact of serious games on science learning achievement compared with more conventional instruction: An overview and a meta-analysis. *Studies in Science Education*, *55*(2), 169–214. https://doi.org/10.1080/03057267.2019.1722420

Robison, W. B. (2013). Stimulation, Not Simulation: An Alternate Approach to History Teaching Games. *The History Teacher*, *46*(4), 577–588.

Rosli, H., Kamaruddin, N., Wirasari, I., & Isa, B. (2022). Understanding the Digital Storytelling

Process for Museum Exhibition through Content Analysis. *Asian Journal of Environment-Behaviour Studies*, *7*(23), Article 23. https://doi.org/10.21834/aje-bs.v7i23.415

Rotgans, J. I., & Schmidt, H. G. (2017). Interest development: Arousing situational interest affects the growth trajectory of individual interest. *Contemporary Educational Psychology*, *49*, 175–184. https://doi.org/10.1016/j.cedpsych.2017.02.003

Sabourin, J. L., & Lester, J. C. (2014). Affect and Engagement in Game-BasedLearning Environments. *IEEE Transactions on Affective Computing*, *5*(1), 45–56. https://doi.org/10.1109/T-AFFC.2013.27

Sakaki, M., Murayama, K., Frenzel, A. C., Goetz, T., Marsh, H. W., Lichtenfeld, S., & Pekrun, R. (2024). Developmental trajectories of achievement emotions in mathematics during adolescence. *Child Development*, *95*(1), 276–295. https://doi.org/10.1111/cdev.13996

Schaper, M.-M., Santos, M., Malinverni, L., Zerbini Berro, J., & Pares, N. (2018). Learning about the past through situatedness, embodied exploration and digital augmentation of cultural heritage sites. *International Journal of Human-Computer Studies*, *114*, 36–50. https://doi.org/10.1016/j.ijhcs.2018.01.003

Schumacher, A., Kammerer, Y., Scharinger, C., Gottschling, S., Hübner, N., Tibus, M., Kasneci, E., Appel, T., Gerjets, P., & Bardach, L. (2025). How do intellectually curious and interested people learn and attain knowledge? A focus on behavioral traces of information seeking. *European Journal of Personality*, 08902070241309124. https://doi.org/10.1177/08902070241309124

Senko, C. (2016). Achievement Goal Theory: A Story of Early Promises, Eventual Discords, and Future Possibilities. In *Handbook of Motivation at School* (2nd ed.). Routledge.

Spring, D. (2015). Gaming history: Computer and video games as historical scholarship. *Rethinking History*, *19*(2), 207–221. https://doi.org/10.1080/13642529.2014.973714

Stiller, K. D., & Schworm, S. (2019). Game-Based Learning of the Structure and Functioning of Body Cells in a Foreign Language: Effects on Motivation, Cognitive Load, and Performance. *Frontiers in Education*, *4*. https://doi.org/10.3389/feduc.2019.00018

Stoel, G. L., van Drie, J. P., & van Boxtel, C. A. M. (2015). Teaching towards historical expertise. Developing a pedagogy for fostering causal reasoning in history. *Journal of Curriculum Studies*, *47*(1), 49–76. https://doi.org/10.1080/00220272.2014.968212

Sweller, J. (2011). CHAPTER TWO - Cognitive Load Theory. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of Learning and Motivation* (Vol. 55, pp. 37–76). Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Taub, M., Sawyer, R., Lester, J., & Azevedo, R. (2020). The Impact of Contextualized Emotions on Self-Regulated Learning and Scientific Reasoning during Learning with a Game-Based Learning Environment. *International Journal of Artificial Intelligence in Education*, *30*(1), 97–120. https://doi.org/10.1007/s40593-019-00191-1

Tze, V. M. C., Daniels, L. M., & Klassen, R. M. (2016). Evaluating the Relationship Between Boredom and Academic Outcomes: A Meta-Analysis. *Educational Psychology Review*, *28*(1), 119–144. https://doi.org/10.1007/s10648-015-9301-y

Ubisoft. (2007). *Assassin's Creed (series)* [Computer software].

van Boxtel, C., & van Drie, J. (2018). Historical Reasoning: The Interplay of Domain-Specific and Domain-General Aspects. In *Scientific Reasoning and Argumentation*. Routledge.

Viccari, A., & Bardach, L. (2025). Designing Limes Text: Adapting an Educational Video Game into a Text-Based Format for Controlled Experimental Comparisons. *European Conference on Games Based Learning*, *19*(2), 855–863. https://doi.org/10.34190/ecgbl.19.2.3972

Viccari, A., Göllner, R., Hahn, J.-U., & Bardach, L. (2024). Limes (the Roman Frontier): Developing a Video Game for History Learning. *European Conference on Games Based Learning*, *18*(1), Article 1. https://doi.org/10.34190/ecgbl.18.1.2833

Wainwright, A. M. (2014). Teaching Historical Theory through Video Games. *History Teacher*, *47*(4), 579–612.

Wineburg, S. (2001). *Historical Thinking & Other Unnatural Acts: Charting the Future of Teaching the Past*.

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, *105*(2), 249–265. https://doi.org/10.1037/a0031311

Yılmaz, E., Yel, S., & Griffiths, M. D. (2022). Comparison of children's social problem-solving skills who play videogames and traditional games: A cross-cultural study. *Computers & Education*, *187*, 104548. https://doi.org/10.1016/j.compedu.2022.104548

# 4.2

# Paper 4

Viccari, A., Lindner, M., Kienitz, A., Hahn, J.-U., Göllner, R., & Bardach, L. (2026) The Sound of the Roman Empire: Effects of Playing a History Videogame With and Without Sound. *Manuscript in revision.*

Preprint can be accessed at https://doi.org/10.31234/osf.io/7ztn4_v2

**Abstract**

*Background*: Sound is a core component of digital games, and its integration is assumed to support learning, motivation, and positive emotions. However, empirical evidence on the role of sound in educational video games remains limited, particularly in narrative-driven educational adventure games such as digital history games.

*Methods*: In a laboratory experiment, university students ($N$ = 111) either played an educational history video game with sound (ambient audio, character voices, and narrated codex entries providing additional historical information) or without sound. Post-test measures assessed factual knowledge, triggered and maintained situational interest, and academic emotions (enjoyment and boredom). Engagement with optional supplemental historical information provided through in-game codex entries was measured using behavioural log data. Group differences were analysed using one-way analyses of covariance, controlling for relevant pre-test variables.

*Results*: Participants in the sound condition did not score significantly higher on the knowledge test than those in the no-sound condition. Likewise, no statistically significant differences emerged in situational interest, enjoyment, boredom, or codex engagement between conditions. Additional analyses indicated that participants' interactions with codex entries positively predicted knowledge test performance, indicating that voluntary engagement with supplemental content contributed to learning.

*Conclusion*: Our findings suggest that the presence of sound alone may not enhance academic outcomes in a narrative-driven educational video game. Additionally, our findings indicate that learning outcomes depended strongly on learners' engagement with in-game codex entries. Overall, our results on the inclusion of sound highlight the importance of examining specific design features within educational history videogames.

*Keywords:* Game-Based Learning, History Learning, Academic Outcomes, Motivation, Academic Emotions, Video Games

**The Sound of the Roman Empire: Effects of Playing a History Video Game With and Without Sound**

Sound is a central component of digital environments and has been assumed to play an important role in shaping learning experiences and outcomes. In multimedia learning research, the inclusion of audio elements is grounded in the Cognitive Theory of Multimedia Learning (CTML; Mayer, 2002), which proposes that people process information through two limited-capacity channels, one visual-pictorial and one auditory-verbal. When instructional materials distribute information across these channels, learning can be supported by reducing overload on any single channel and facilitating more efficient working-memory use. Interactive computer-based learning environments, including video games, are well-suited to leverage multimedia learning principles, as they often combine visual and auditory information within interactive tasks (Plass et al., 2015; Sella et al., 2016; Wouters & van Oostendorp, 2017). In line with this, a key principle derived from CTML is the modality principle, which shows that narration can enhance understanding and retention compared to on-screen text, particularly when visual information is complex (Mayer & Moreno, 2003; Moreno & Mayer, 1999). According to Mayer's multimedia learning framework, coordinated narration and visual information should improve retention and transfer, highlighting how audio can support coherent mental model formation (Mayer, 2002), which may, in turn, facilitate *knowledge acquisition* in digital game-based learning (Wouters et al., 2013).

In addition to cognitive processing, sound has been proposed to influence *affective* and *motivational aspects of learning*. The present study focuses on academic emotions (i.e., enjoyment, boredom (Pekrun, 2006; Pekrun & Perry, 2014)) and situational interest (Hidi & Renninger, 2006). However, since these outcomes have rarely been examined in research on video games and sound effects, we next review the broader literature on affective and motivational outcomes. Audio elements—such as background music, environmental sound, and character voices—are integral design features that can affect emotional responses, sense of immersion, and engagement. For instance, musical scores can intensify emotional involvement (Lipscomb & Zehnder, 2004), while the presence of game sound has been shown to enhance experiential states such as immersion, flow, positive affect, and perceived competence in interactive environments (Nacke et al., 2010). In game-based learning environments, the presence of audio narration or voiceovers can enhance behavioural engagement, as demonstrated by findings that learners exhibited higher involvement when game sound was present (Byun & Loh, 2015). It has also been suggested that music may influence immersion and perceived time flow (Sanders & Cairns, 2010). At the same time, research incorporating

music as a design element in educational games indicates that its effects are not uniformly positive across contexts. For example, studies that embed background music within broader instructional or design frameworks—such as those focusing on learner characteristics or system-level customization—report heterogeneous motivational and performance-related outcomes (Ku et al., 2016; Yang et al., 2019). These findings suggest that the impact of game audio may depend on how it is implemented and integrated, rather than on its mere presence. Furthermore, it has been shown that adding a self-similar avatar voice enhanced learners' enjoyment and perceived competence and influenced their in-game performance (Kao et al., 2021). Although this manipulation involved personalized voice output rather than ambient sound design, it demonstrates that audio features embedded in educational games may meaningfully alter emotional and motivational outcomes.

Sound also plays a role in directing attention, focus, and information selection, and shapes actual behaviours in interactive environments (Byun & Loh, 2015; Moreno & Mayer, 1999). Specifically, the signalling principle as deduced from CTML proposes that audio cues can help learners direct their attention to important information (Mayer & Moreno, 2003). Conceptual work on audio game design likewise emphasizes that sound can structure players' spatial orientation, guide navigation, and support attentional focus within interactive environments (Rovithis et al., 2014). From this perspective, audio does not merely accompany visual information but functions as an informational layer that scaffolds perception, facilitates navigation, and shapes how players interact with the game world. Thus, the inclusion of sound may also influence learners' in-game behaviours, such as engagement with optional or supplemental content, by enhancing immersion and supporting focused engagement.

Despite strong theoretical arguments and some promising empirical insights, research on the role of sound in educational video games remains limited. Whereas many studies focus on visual design, narrative structure, or gameplay mechanics, comparatively few have examined whether audio elements affect knowledge outcomes, motivation, emotional experiences, or in-game behaviours. This gap may be particularly relevant for subjects such as history: For history learning, immersion, emotional engagement, and the construction of coherent mental models are central to how learners make sense of historical narratives and environments (McCall, 2023). Narrative-oriented analyses further suggest that sound is a key factor in establishing coherent fictional story worlds and fostering immersion in narrative game environments (Domsch, 2016). Such immersive narrative experiences are often considered relevant to learner engagement in educational history games (McCall, 2016). Accordingly, empirical research

systematically investigating the role of sound in educational history games is needed to gain deeper insights into audio design elements that may facilitate learners' knowledge acquisition, interest, emotional responses, and engagement with historical content.

## The Present Study

The present study examined the effects of sound on knowledge test performance, motivational outcomes (triggered and maintained situational interest), academic emotions (enjoyment and boredom), and in-game behaviour (engagement with codex entries providing historical information) in *Limes* (Viccari et al., 2024), an educational video game on the Roman Empire. In a between-subjects design, participants were randomly assigned to one of two conditions: a version of the game with sound or a version without sound. All learners engaged with the game for up to 90 minutes and subsequently completed the post-test measures. All hypotheses, measures, covariates, exclusion rules, and statistical analyses were preregistered prior to conducting the analyses (blinded link for peer-review: https://aspredicted.org/xa6gt7.pdf).

Based on the theoretical frameworks and empirical findings reviewed above, we specified four main hypotheses. First, we hypothesized that participants in the sound condition would achieve higher scores on the post-test knowledge measure than those in the no-sound condition, as the inclusion of audio narration and ambient sound may support cognitive processing and the formation of coherent mental models (R. E. Mayer, 2002) (*Hypothesis 1, H1*).

Second, we expected participants exposed to sound to report higher levels of situational interest (*Hypothesis 2, H2*). Although prior research on sound effects in educational video games that directly measures situational interest is, to the best of our knowledge, lacking, there is some evidence on effects of background music and narrative elements on learners' motivation (Kao et al., 2021; Lipscomb & Zehnder, 2004). Drawing on the interest development framework proposed by Hidi and Renninger (2006), sound effects are expected to trigger situational interest by increasing the perceptual salience and experiential vividness of the learning environment. Insofar as sound effects help sustain engagement with task-relevant information across the learning episode, they may also support maintained situational interest.

Third, building on prior research showing positive effects on learners' emotional experiences (Lipscomb & Zehnder, 2004; Nacke et al., 2010), we hypothesized that learners in the sound condition would report more adaptive academic emotions—specifically, higher

enjoyment and lower boredom—compared with those in the no-sound condition (*Hypothesis 3, H3*).

Finally, we expected that participants in the sound condition would more often explore additional historical information (referred to in the game as codex entries) and for a longer duration (*Hypothesis 4, H4*), reflecting the role of sound in guiding attention and information selection within interactive learning environments (Byun & Loh, 2015; Moreno & Mayer, 1999).

## Methods

### Participants and Design

A total of 112 university students enrolled in psychology, teacher education, and sport science programmes participated in the study (84 women, 28 men; $M_{age}$= 23.19, *SD* = 3.95). Data collection was carried out in a controlled laboratory setting at a German university. Based on an a priori power analysis using G*Power 3.1 (Faul et al., 2009) for an analysis of covariance comparing two independent groups, we aimed to recruit at least N = 102 participants, assuming a medium effect size ($f$ = 0.25), an alpha level of .05, and a power of .80. To account for potential exclusions, we collected data from 112 individuals. The study received ethics approval from the University of [blinded for peer-review] ethics board (approval code: [blinded]). Figure 3 presents an overview of the study procedure.

Participants were recruited through multiple channels, including campus flyers, university mailing lists, and the university's participation platform, where the study was listed as a standard in-lab study offering course credit as compensation. To be eligible, participants had to be at least 18 years old, fluent in German, not enrolled in a history degree programme, and not have uncorrected severe visual impairments.

Participants were randomly assigned to play the video game either with or without sound. We collected data from 76 participants and included 36 participants from a previous study for the no-sound condition. The previous study also employed random assignment to different conditions and focused on a different research question (a comparison of the *Limes* game and a parallel text format, and of two instructions: having fun versus learning), but did not involve sound (Viccari et al., 2025). We included only participants from the game condition of the previous study who had received the learning instruction to align with the participant instruction in the current study. Overall, 56 participants (40 women, 16 men) were in the *sound* condition, and 56 participants (44 women, 12 men) were in the *no-sound* condition.

**Experimental Manipulation and Materials**

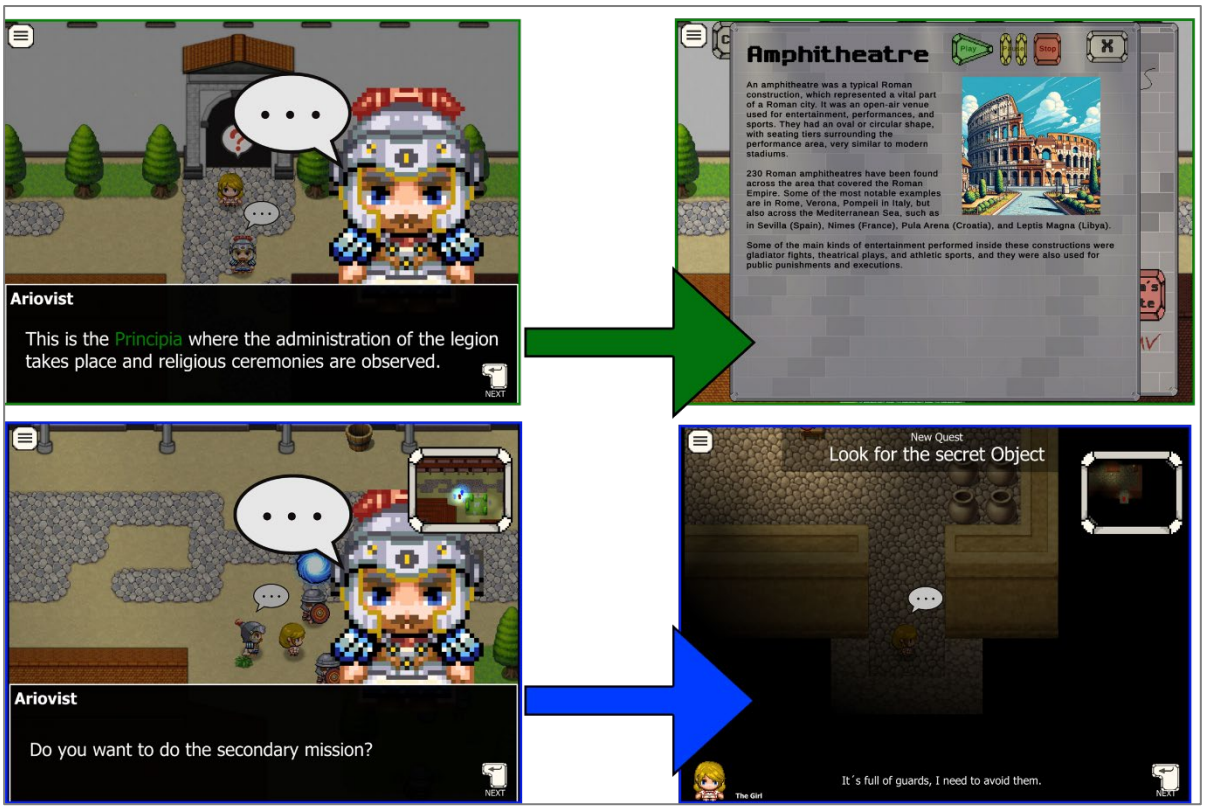
In our between-subjects design (sound versus no sound), participants engaged with Limes, an educational 2D top-down adventure game developed in Unity (version 2021.3.7f1) (Viccari et al., 2024).

In the sound condition, ambient sounds, character voices, and codex entries audios were all enabled. In the no-sound condition, these audio elements were disabled, while all other game features remained identical. This manipulation allowed us to isolate the role of audio cues on engagement and learning. Prior to engaging with *Limes*, all participants received the instruction that they should use the video game to learn as much as possible about the history content.

*Limes* presents a historically grounded narrative set along the Roman frontier during the third century CE. Players follow the journey of a Germanic child who is taken by Roman soldiers and gradually forms an unexpected friendship with a Roman centurion, Ariovist. The story unfolds across five levels, each containing a main storyline, optional secondary missions, and supplementary codex entries that provide additional historical context (see Figure 1). Codex entries (see Figure 2 for an example) are short and contain texts covering topics such as Roman emperors, military structures, and everyday life along the limes (the Roman Empire's border).

**Figure 1**

*Illustration of codex entry access (top) and secondary mission activation (bottom) in the Limes video game*

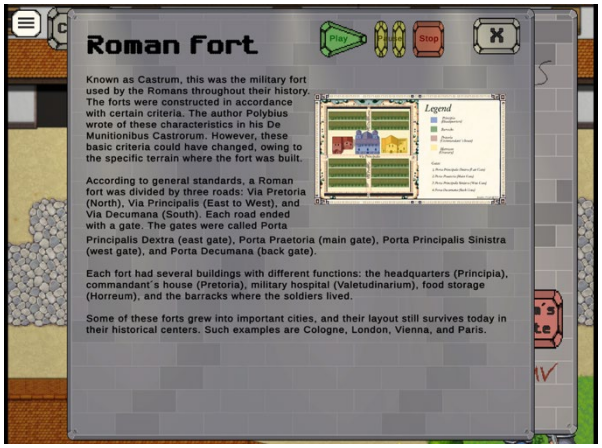



*Note.* The top panel shows how players open a codex entry after interacting with a character. The bottom panel shows the transition from dialogue into a secondary mission, where the player enters an interactive gameplay sequence.

The game includes 36 codex entries distributed across the five levels. Players can voluntarily open these entries by interacting with characters or objects in the game world. Each entry consists of a short text; in the German version used in this study, entries averaged 148 words (range 59–265), totalling 6,328 words of codex material, and were unlocked through interactions with in-game characters or objects. Across all participants, behaviours were automatically recorded, including time spent reading codex entries, exploration of secondary missions, and overall time-on-task (see Table 1 for descriptive engagement measures).

**Figure 2**

*Codex entry as shown in the video game*




Before beginning the task, participants viewed a short instruction screen that explained the study procedure and encouraged them to engage thoroughly with the available content. This screen remained visible until the experiment leader manually launched the game, and the time spent on the instruction screen was recorded.

All content was designed and reviewed with the aim of providing accurate historical material in an interactive format appropriate for university students. A detailed description of the narrative design, structural elements, and development process of *Limes* can be found in Viccari et al. (2024).

## Measures

**Pre-Test Background Variables and Covariates**. In the pre-test, we assessed several background variables as potential covariates for the analyses. Prior knowledge of the learning content was assessed using a short 5-item version of the knowledge test administered at pre-test (possible range: 0–5). This score was included as a covariate in subsequent analyses. Prior interest in history was measured using three items adapted from Schumacher et al. (2025) (e.g., "How interested are you in Roman history?", rated on a 7-point Likert scale ($\alpha$ = .73). Prior self-reported historical knowledge ("How much knowledge do you have about the following topics?"), was measured with three items, each rated on a 7-point Likert scale (1 = no knowledge at all, 7 = a great deal of knowledge). The items assessed self-reported knowledge of Roman, ancient, and general history (adapted from Schumacher et al. (2025)). Responses were averaged to form a composite scale score ($\alpha$ = .78). Participants also reported their average weekly gaming hours.

**Knowledge Test.** The knowledge test assessed factual knowledge of Roman history, including topics such as Roman society, military structures, frontier life along the Roman border (limes), and key historical events and figures covered in the learning materials. It consisted of 25 single-choice questions (with 4 answer options, plus an "I do not know" option) that were coded as correct or incorrect. A total sum score was calculated based on the number of correctly answered items (possible range: 0–25). Internal consistency was high ($\alpha_{sound}$ = .79, $\alpha_{no\text{-}sound}$= .81).

**Situational Interest**. Situational interest was measured using eleven items adapted from Linnenbrink-Garcia et al. (2010). The scale included three components: triggered situational interest (4 items; e.g., "Playing *Limes* is exciting"), maintained interest feeling (4 items; e.g., "I am enthusiastic about what I learn in *Limes*"), and maintained interest value (3 items; e.g., "What I learn in Limes is valuable to me"). All items were rated on a 7-point scale (1 = does not apply at all, 7 = fully applies). Following Krebs et al. (2024), the two maintained interest components were combined into a single maintained-interest scale for further analysis. Internal consistency was high for both triggered interest ($\alpha$ = .86 overall; $\alpha_{sound}$ = 85; $\alpha_{no\text{-}sound}$= .88) and maintained interest ($\alpha$ = .89 across all participants; $\alpha_{sound}$ = 87; $\alpha_{no\text{-}sound}$= .91).

**Academic Emotions.** Academic emotions were assessed using the short version of the Achievement Emotions Questionnaire (AEQ-S; Bieleke et al., 2021). The measure consists of four items assessing enjoyment (e.g., "Engaging with the content of *Limes* brings me joy") and four items assessing boredom (e.g., "The content of Limes is so boring that I catch myself daydreaming"). Participants responded on a 5-point Likert scale (1 = do not agree at all, 5 = fully agree). Internal consistency was acceptable to high for enjoyment ($\alpha$ = .80 overall, $\alpha_{sound}$ = 73; $\alpha_{no\text{-}sound}$ = .86), and boredom ($\alpha$ = .82 overall, $\alpha_{sound}$ = 73; $\alpha_{no\text{-}sound}$ = .87). All measures used in our study are reported in the Online Supplement.

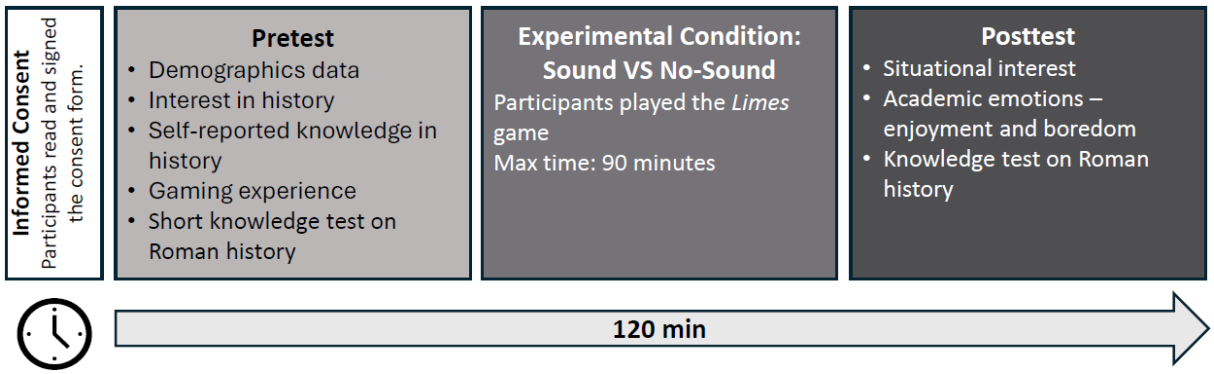
## Procedure

The experiment took place in a laboratory setting at a German university. Upon arrival, participants were welcomed by the experiment leader, informed about the study procedures and data protection, and they provided informed consent. Participants then completed the pre-test on Unipark via Firefox on university-provided computers. The pre-test included demographic information, self-reported history knowledge, self-reported interest in history, and a short knowledge test on Roman history. After the pre-test, participants received a short instruction screen explaining the task and encouraging them to engage thoroughly with the game content.

Assignment to the audio condition (sound vs. no-sound) was determined by randomization embedded in the desktop applications: in the sound condition, ambient audio, character voices, and voiceover of the codex entry were enabled; in the no-sound condition, these audio elements were disabled while all other features remained identical. Once participants had read the instruction screen, the experiment leader manually launched the game application for them. Participants were instructed to spend a maximum of 90 minutes engaging with the game and wore headphones during the session. Time spent on the instruction screen and all in-game behaviours (e.g., codex engagement, secondary missions, total playtime) were automatically recorded. After the learning phase—or once the allocated time had expired—participants returned to the browser to complete the post-test on Unipark, which included measures of situational interest, academic emotions (enjoyment and boredom), and the 25-item factual knowledge test on Roman history. Each laboratory session lasted up to 120 minutes in total.

**Figure 3**

*Overview of the Experimental Procedure*




*Note*. Maximum completion time maximum was 120 min. In-game behavioural data (time in the codex, codex clicks) was collected while participants played Limes.

**Analyses**

We conducted one-way analyses of covariance (ANCOVAs) to examine the effect of the audio condition (sound vs. no-sound) on all outcome variables. Correlations between pre-test variables and dependent variables were inspected to determine their inclusion as covariates. All significance tests were evaluated at the .05 level, and all directional hypotheses were assessed using one-tailed tests.

## Results

### Preliminary Analyses

As preregistered, participants were to be excluded if their total time spent engaging with Limes fell below 20 minutes. No participant met this exclusion criterion. One participant was excluded due to a procedural deviation that prevented correct data recording, resulting in a final sample of $N = 111$ for all analyses. On average, participants spent 1.63 minutes on the instruction screen, followed by approximately 75 minutes interacting with the game material, and an additional 9.3 minutes completing the post-test (see Table 1).

To examine whether the two sound conditions differed on pre-test variables, we conducted t-tests, which showed no significant differences between the sound and no-sound groups in age $t(108.10) = -0.78$, $p = .44$, weekly gaming hours, $t(69.18) = 0.22$, $p = .83$, prior historical knowledge, $t(108.41) = -0.14$, $p = .89$, interest in history, $t(107.21) = -0.41$, $p = .68$, or pretest factual knowledge, $t(108.95) = -0.08$, $p = .93$. Chi² tests likewise indicated no difference in gender distribution across conditions ($\chi^2(1) = 0.51$, $p = .477$) or gaming experience, ($\chi^2(1) = 0.54$, $p = .463$). These results suggest that the two experimental groups were comparable on the pre-test variables assessed.

**Table 1**

*Time spent in Limes Game (in Minutes) by Condition*

| | No-Sound ($n = 56$) | | Sound ($n = 55$) | |
|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* |
| Instruction Screen Time (Before Game) | 1.62 | 0.46 | 1.64 | 0.29 |
| Codex Entry Time in Game | 9.82 | 9.54 | 10.15 | 8.16 |
| Game Play Time | 64.03 | 12.96 | 65.85 | 11.66 |
| Total Time Spent in Game | 73.85 | 13.73 | 76.00 | 11.56 |
| Post-Test Total Time | 9.35 | 2.29 | 9.32 | 1.62 |

**Table 2**

*Means and Standard Deviations for Outcome Variables*

| | No-Sound ($n$ = 56) | | Sound ($n$ = 55) | |
|---|---|---|---|---|
| Outcomes Variables | *M* | *SD* | *M* | *SD* |
| Post-Test Knowledge (Score Range: 0–25) | 13.95 | 4.89 | 13.80 | 4.59 |
| Triggered Situational Interest (Scale: 1–7) | 4.56 | 1.26 | 4.69 | 1.25 |
| Maintained Situational Interest (Scale: 1–7) | 4.28 | 1.19 | 4.49 | 1.04 |
| Enjoyment (Scale: 1–5) | 3.39 | 0.86 | 3.47 | 0.69 |
| Boredom (Scale: 1–5) | 1.73 | 0.78 | 1.63 | 0.59 |
| Codex Clicks (Count) | 15.79 | 13.34 | 18.53 | 11.86 |
| Time in Codex (Minutes) | 9.82 | 9.54 | 10.15 | 8.16 |

To determine whether pre-test characteristics should be controlled for in the main analyses, we inspected the correlations between all pre-test measures and the dependent variables (see Table 3). In this screening, several pre-test variables showed meaningful associations with the outcomes. Accordingly, interest in history, self-reported prior knowledge, and pre-test factual knowledge were included as covariates only in those ANCOVA models in which they showed a significant association with the respective dependent variable.

**Table 3**

*Overview of Correlations Between Study Variables*

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Interest in History (Pre-test) | 1.00 | | | | | | | | | | |
| 2. Self-Reported Knowledge (Pre-test) | .65** | 1.00 | | | | | | | | | |
| 3. Pre-Test Knowledge (5 Items) | .41*** | .54*** | 1.00 | | | | | | | | |
| 4. Weekly Gaming Hours (Pre-test) | .13 | .14 | .03 | 1.00 | | | | | | | |
| 5. Post-Test Knowledge Test (25 Items) | .33** | .31** | .35** | -.09 | 1.00 | | | | | | |
| 6. Triggered Situational Interest | .27* | .14 | -.02 | .08 | .28* | 1.00 | | | | | |
| 7. Maintained Situational Interest | .52*** | .36** | .22 | .04 | .51*** | .73*** | 1.00 | | | | |
| 8. Enjoyment | .42*** | .31** | .13 | .04 | .36** | .84*** | .83*** | 1.00 | | | |
| 9. Boredom | -.20 | -.15 | -009 | .03 | -.30** | -.68*** | -.44*** | -.56*** | 1.00 | | |
| 10. Codex Clicks | .08 | .07 | .06 | -.09 | .64*** | .24* | .39*** | .33** | -.25* | 1.00 | |
| 11. Time in Codex (Seconds) | .06 | .02 | .04 | -.14 | .61*** | .33** | .42*** | .35** | -.25* | .85*** | 1.00 |

*Note*. *Correlation is significant at the .05 level (two-tailed); **Correlation is significant at the .01 level (two-tailed); ***Correlation is significant at the .001 level (two-tailed).

## Hypothesis Testing

**Effects on Situational Interest (Hypothesis 1)**. The ANCOVA for *triggered situational interest* yielded no significant effect of sound, $F(1,108) = 0.22$, $p = .639$, $\eta^2_p < .01$. Participants in the sound condition reported triggered interest similar to those in the no-sound condition (Sound: $M = 4.69$, $SD = 1.25$; No-Sound: $M = 4.56$, $SD = 1.26$). Pre-test interest in history was a significant covariate, $F(1,108) = 6.42$, $p = .013$, $\eta^2_p = .06$. For *maintained situational interest*, the ANCOVA again revealed no significant effect of sound, $F(1,107) = 0.85$, $p = .360$, $\eta^2_p = .01$. Means were similar between the sound condition ($M = 4.49$, $SD = 1.04$) and the no-sound condition ($M = 4.28$, $SD = 1.19$). Pre-test interest in history significantly predicted maintained interest, $F(1,107) = 20.65$, $p < .001$, $\eta^2_p = .26$.

**Effects on Knowledge Acquisition (Hypothesis 2).** The ANCOVA on the 25-item *knowledge post-test* revealed no significant effect of sound on students' knowledge acquisition during the game, $F(1,106) = 0.10$, $p = .752$, $\eta^2_p < .01$. Participants in the sound condition ($M = 13.78$, $SD = 4.59$) and the no-sound condition ($M = 13.86$, $SD = 4.89$) achieved comparable scores. Pre-test knowledge, $F(1,106) = 8.78$, $p = .004$, $\eta^2_p = .16$, and pre-test interest in history,

$F(1,106) = 4.72$, $p = .032$, $\eta^2_p = .06$, significantly predicted post-test knowledge scores.

**Effects on Academic Emotions (Hypothesis 3).** The ANCOVA on *enjoyment* showed no significant effect of sound, $F(1,107) = 0.15$, $p = .698$, $\eta^2_p < .01$. Enjoyment ratings were comparable across conditions (Sound: $M = 3.47$, $SD = 0.69$; No-Sound: $M = 3.39$, $SD = 0.86$). Pretest interest in history emerged once again as a significant covariate, $F(1,107) = 7.52$, $p = .007$, $\eta^2_p = .11$. Finally, the ANCOVA for boredom indicated no significant effect of sound, $F(1,108) = 0.43$, $p = .513$, $\eta^2_p < .01$. Participants in the sound condition reported similar boredom levels ($M = 1.63$, $SD = 0.59$) as those in the no-sound condition ($M = 1.73$, $SD = 0.78$).

**Effects on Codex Engagement (Hypothesis 4).** The ANCOVA on the number of codex clicks revealed no significant effect of sound, $F(1,108) = 1.28$, $p = .260$, $\eta^2_p = .01$. Participants in the sound condition opened codex entries at similar rates ($M = 18.51$, $SD = 11.91$) as those in the no-sound condition ($M = 15.77$, $SD = 13.26$). Self-reported prior historical knowledge did not significantly predict codex access, $F(1,108) = 0.25$, $p = .616$. A second ANCOVA examined codex engagement time. Again, there was no significant effect of sound, $F(1,108) = 0.04$, $p = .844$, $\eta^2_p < .01$. Learners in the sound condition spent comparable time in the codex ($M = 10.24$ min, $SD = 8.16$) as those in the no-sound condition ($M = 9.82$ min, $SD = 9.54$). Self-reported prior historical knowledge did not significantly predict codex engagement time, $F(1,108) = 0.01$, $p = .927$.

***Additional analyses (not preregistered)***

**Bayesian follow-up analyses**. To further evaluate the strength of evidence for the null versus the alternative hypothesis (see Hypotheses 1-4), we conducted Bayesian independent-samples t-tests for all main outcome variables. The analyses provided moderate evidence in favour of the null hypothesis for time in the codex ($BF_{01} = 3.63$), triggered situational interest ($BF_{01} = 3.25$), enjoyment ($BF_{01} = 3.32$), and boredom ($BF_{01} = 2.97$). Evidence for maintained situational interest ($BF_{01} = 2.43$) and codex clicks ($BF_{01} = 2.15$) was weak and favoured the null hypothesis. For post-test knowledge scores, the Bayes factor was inconclusive ($BF_{01} = 0.79$), meaning the evidence was insufficient to support either the null or the alternative hypothesis. Accordingly, conclusions about the effect of sound on knowledge test performance remain tentative. Overall, the Bayesian results converge with the null effects observed in the primary analyses, suggesting that the presence of sound did not have a meaningful impact on learners' interest, emotions, and knowledge test scores.

**Effects of Codex Engagement**. Following the observation of strong correlations

between codex engagement measures and post-test knowledge scores (see Table 3), we conducted two separate multiple regression analyses predicting knowledge test performance from codex clicks and codex viewing time, respectively, while controlling for relevant covariates. In the first model, codex clicks significantly predicted knowledge scores ($b = 0.21$, $p < .001$), and in the second model, time in the codex significantly predicted knowledge scores ($b = 0.30$, $p < .001$). Prior knowledge was also a significant predictor in both models ($ps < .001$).

## Discussion

The present study examined how sound in an educational history video game affects knowledge test performance, motivation, academic emotions, and beneficial in-game behaviours (engagement with the codex entries). We compared two versions of the same history video game that differed only in the sound condition, including ambient audio, character voices, and codex narration. Below, we discuss the main findings for each hypothesis, as well as the exploratory findings.

Sound was expected to support learning by distributing processing across auditory and visual channels, thereby reducing visual overload (Mayer, 2002). Contrary to our preregistered hypothesis, however, the inclusion of sound did not lead to higher factual knowledge as assessed by the post-test. Replications of our work are needed before definite conclusions can be drawn; however, we propose that one potential explanation is that Limes presents a largely static, text-based informational structure, in which the cognitive load imposed by visual materials may not have been sufficiently high for audio narration to offer an advantage. Prior CTML research shows that modality benefits are most pronounced when learners process fast-paced animations or visually dense demonstrations (Mayer & Moreno, 2003; Moreno & Mayer, 1999), whereas narration offers limited benefits when learners have ample time to process written information. The structure of Limes, particularly its self-paced reading segments and extensive textual codex entries, may therefore not have placed learners in conditions where audio would meaningfully lighten cognitive demands.

In our study, sound did not significantly affect situational interest or academic emotions. This stands in contrast to findings indicating that music and sound effects can enhance immersion, affect, and behavioural engagement (Byun & Loh, 2015; Lipscomb & Zehnder, 2004; Nacke et al., 2010). As a first potential explanation, the overall very low levels of boredom observed in our study suggest that the game was already perceived as sufficiently engaging, even without audio. The reward of engaging with the game may therefore not have

been low in the no-sound condition, leaving limited scope for additional increases in positive activation through sound (Bardach & Murayama, 2025; Eitel et al., 2025). Another potential explanation is that the audio elements in Limes—ambient sounds and sparse voiceovers—may have been relatively subtle compared to continuous, event-linked audio designs often used in studies that reported motivational benefits (Nacke et al., 2010; Sanders & Cairns, 2010). A further explanation comes from research showing that not all types of audio exert the same motivational influence in educational games (Erhel & Jamet, 2013; Kao et al., 2021). For example, Kao et al. (2021) found that audio can have meaningful motivational effects when it enhances self-relevance and identification with the avatar rather than merely providing background stimulation. Thus, the audio used in future experiments on the effects of voice in history video games may benefit from personalised, identity-linked implementations.

Exploration of codex entries did not differ between the sound and no-sound conditions. This may indicate that codex use in Limes is not driven by peripheral sensory features such as ambient audio, but could instead reflect learners' intrinsic motivation and perceived relevance of the task in the game. Research on curiosity and interest supports this pattern: exploratory behaviours are typically initiated when learners perceive informational value or anticipate epistemic rewards, rather than through sensory features alone (Murayama et al., 2019). Similarly, models of situational interest suggest that whereas environmental features may trigger initial interest, sustained engagement depends primarily on learners' enjoyment of and perceived value in the content (Linnenbrink-Garcia et al., 2010). In this context, codex entries provide substantive historical information, and learners may engage with them when they perceive this information as useful for understanding the game world or succeeding in tasks. Interestingly, although we did not assess learners' motivation during gameplay, the bivariate correlations between codex measures and our motivational and emotional outcomes are all positive and moderate in size (see Table 3).

Moreover, additional analyses indicated that codex engagement positively predicted knowledge test performance, even after controlling for prior knowledge. These findings reinforce the instructional value of the codex system and highlight that voluntary engagement with supplemental content appears to be a central mechanism driving learning in Limes, regardless of whether sound is present.

**Limitations and Directions for Future Research**

Several limitations should be acknowledged. First, the sound manipulation may have

been too subtle to produce measurable differences. Ambient audio and character voices offer a different experiential profile than continuous narration or directive audio cues, which may be more likely to influence cognitive outcomes or motivational states. Second, the study manipulated the presence versus absence of sound but did not vary the type, intensity, or instructional function of audio. Future research could compare ambient sound, music, fully narrated content, and signalling cues to determine which audio forms offer the strongest learning benefits. Third, the laboratory setting and the use of a single instructional approach may have limited learners' natural engagement with audio. Research on instructional goals suggests that the framing of educational games can influence learners' processing of game content and learning outcomes (Erhel & Jamet, 2019). Studies using different task instructions may therefore reveal effects that were not observable under the single instructional framing used in the present study.

Additionally, while codex engagement predicted knowledge test performance, the codex entries consisted primarily of written text. Although audio playback was available for every entry, the current study did not track whether participants used this feature. As a result, we cannot determine whether codex audio contributed to learning or whether it was simply underutilized. Future work could develop narrated or partially narrated codex entries with a clearer instructional purpose and examine how different audio formats influence learners who struggle with dense textual material. Adaptive audio support, guided by principles of multimedia learning, may also offer opportunities to enhance comprehension and engagement (Moreno & Mayer, 1999).

## Conclusions

The present study provides a systematic examination of the role of sound in Limes, an educational history video game. Across knowledge, situational interest, and academic emotions, sound did not yield measurable benefits, nor did it alter learners' use of supplemental content, as measured by codex engagement. These findings contribute to our understanding of when and whether sound matters in game-based learning.

**References**

Bardach, L., & Murayama, K. (2025). The role of rewards in motivation—Beyond dichotomies. *Learning and Instruction*, *96*, 102056. https://doi.org/10.1016/j.learninstruc.2024.102056

Bieleke, M., Gogol, K., Goetz, T., Daniels, L., & Pekrun, R. (2021). The AEQ-S: A short version of the Achievement Emotions Questionnaire. *Contemporary Educational Psychology*, *65*, 101940. https://doi.org/10.1016/j.cedpsych.2020.101940

Byun, J., & Loh, C. S. (2015). Audial engagement: Effects of game sound on learner engagement in digital game-based learning environments. *Computers in Human Behavior*, *46*, 129–138. https://doi.org/10.1016/j.chb.2014.12.052

Domsch, S. (2016). Hearing Storyworlds: How Video Games Use Sound to Convey Narrative. In J. Mildorf & T. Kinzel (Eds.), *Audionarratology* (pp. 185–198). De Gruyter. https://doi.org/10.1515/9783110472752-012

Eitel, A., Krebs, M.-C., & Schöne, C. (2025). Introducing the FoRe-Squares Model: Focus Support and Reward as Key Predictors of Effective Technology-Augmented Instruction. *Educational Psychology Review*, *37*(3), 67. https://doi.org/10.1007/s10648-025-10045-z

Erhel, S., & Jamet, E. (2013). Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & Education*, *67*, 156–167. https://doi.org/10.1016/j.compedu.2013.02.019

Erhel, S., & Jamet, E. (2019). Improving instructions in educational computer games: Exploring the relations between goal specificity, flow experience and learning outcomes. *Computers in Human Behavior*, *91*, 106–114. https://doi.org/10.1016/j.chb.2018.09.020

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Hidi, S., & Renninger, K. A. (2006). The Four-Phase Model of Interest Development. *Educational Psychologist*, *41*(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

Kao, D., Ratan, R., Mousas, C., & Magana, A. J. (2021). The Effects of a Self-Similar Avatar Voice in Educational Games. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CHI PLAY), 1–28. https://doi.org/10.1145/3474665

Krebs, M.-C., Braschoß, K., & Eitel, A. (2024). Does watching an explainer video help learning with subsequent text? – Only when prompt-questions are provided. *Learning and Instruction*, *94*, 101988. https://doi.org/10.1016/j.learninstruc.2024.101988

Ku, O., Hou, C.-C., & Chen, S. Y. (2016). Incorporating customization and personalization into game-based learning: A cognitive style perspective. *Computers in Human Behavior*, *65*, 359–368. https://doi.org/10.1016/j.chb.2016.08.040

Linnenbrink-Garcia, L., Durik, A. M., Conley, A. M., Barron, K. E., Tauer, J. M., Karabenick, S. A., & Harackiewicz, J. M. (2010). Measuring Situational Interest in Academic Domains. *Educational and Psychological Measurement*, *70*(4), 647–671. https://doi.org/10.1177/0013164409355699

Lipscomb, S. D., & Zehnder, S. M. (2004). Immersion in the Virtual Environment: The Effect of a Musical Score on the Video Gaming Experience. *Journal of PHYSIOLOGICAL ANTHROPOLOGY and Applied Human Science*, *23*(6), 337–343. https://doi.org/10.2114/jpa.23.337

Mayer, R. E. (2002). Multimedia learning. In *Psychology of Learning and Motivation* (Vol. 41, pp. 85–139). Academic Press. https://doi.org/10.1016/S0079-7421(02)80005-6

Mayer, R. E., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, *38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

McCall, J. (2016). Teaching History With Digital Historical Games: An Introduction to the Field and Best Practices. *Simulation & Gaming*, *47*(4), 517–542. https://doi.org/10.1177/1046878116646693

McCall, J. (2023). *Gaming the Past: Using Video Games to Teach Secondary History* (2nd ed.). Routledge.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, *91*(2), 358–368. https://doi.org/10.1037/0022-0663.91.2.358

Murayama, K., FitzGibbon, L., & Sakaki, M. (2019). Process Account of Curiosity and Interest:

A Reward-Learning Perspective. *Educational Psychology Review*, *31*(4), 875–895. https://doi.org/10.1007/s10648-019-09499-9

Nacke, L. E., Grimshaw, M. N., & Lindley, C. A. (2010). More than a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game. *Interacting with Computers*, *22*(5), 336–343. https://doi.org/10.1016/j.intcom.2010.04.005

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

Pekrun, R., & Perry, R. P. (2014). Control-value theory of achievement emotions. In *International handbook of emotions in education* (pp. 120–141). Routledge/Taylor & Francis Group.

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of Game-Based Learning. *Educational Psychologist*, *50*(4), 258–283. https://doi.org/10.1080/00461520.2015.1122533

Rovithis, E., Floros, A., Mniestris, A., & Grigoriou, N. (2014). Audio games as educational tools: Design principles and examples. *2014 IEEE Games Media Entertainment*, 1–8. https://doi.org/10.1109/GEM.2014.7048083

Sanders, T., & Cairns, P. (2010, September 1). *Time perception, immersion and music in videogames*. Proceedings of HCI 2010. https://doi.org/10.14236/ewic/HCI2010.21

Schumacher, A., Kammerer, Y., Scharinger, C., Gottschling, S., Hübner, N., Tibus, M., Kasneci, E., Appel, T., Gerjets, P., & Bardach, L. (2025). How do intellectually curious and interested people learn and attain knowledge? A focus on behavioral traces of information seeking. *European Journal of Personality*, 08902070241309124. https://doi.org/10.1177/08902070241309124

Sella, F., Tressoldi, P., Lucangeli, D., & Zorzi, M. (2016). Training numerical skills with the adaptive videogame “The Number Race”: A randomized controlled trial on preschoolers. *Trends in Neuroscience and Education*, *5*(1), 20–29. https://doi.org/10.1016/j.tine.2016.02.002

Viccari, A., Göllner, R., Hahn, J.-U., & Bardach, L. (2024). Limes (the Roman Frontier): Developing a Video Game for History Learning. *European Conference on Games Based Learning*, *18*(1), Article 1. https://doi.org/10.34190/ecgbl.18.1.2833

Viccari, A., Kienitz, A., Eitel, A., Hahn, J.-U., Göllner, R., Lindner, M., & Bardach, L. (2025). *Exploring Roman History through Video Game and Text: Effects of Format and Instructional Framing* (Fa28t_v1). PsyArXiv. https://doi.org/10.31234/osf.io/fa28t_v1

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, *105*(2), 249–265. https://doi.org/10.1037/a0031311

Wouters, P., & van Oostendorp, H. (2017). Overview of Instructional Techniques to Facilitate Learning and Motivation of Serious Games. In P. Wouters & H. van Oostendorp (Eds.), *Instructional Techniques to Facilitate Learning and Motivation of Serious Games* (pp. 1–16). Springer International Publishing. https://doi.org/10.1007/978-3-319-39298-1_1

Yang, T. C., Chen, M. C., & Chen, S. Y. (2019). The Effects of Background Music on Game-Based Learning: A Cognitive Style Approach. *Asia-Pacific Education Researcher*, *28*(6), 495–508. https://doi.org/10.1007/s40299-019-00450-8

# 5

# SYSTEMATIC REVIEW: PAPER 5

Viccari, A., Fahrbach, T., Schwarzenthal, M., Kienitz, A., Emslander, V., & Bardach, L. (2026). Video Games, Prejudice, and Intercultural Competence: A Systematic Review. *Submitted manuscript.*

Preprint can be accessed at https://doi.org/10.31234/osf.io/m8q73_v1

**Abstract**

**Background**: Digital games are promising tools to support intercultural learning and prejudice reduction. Yet insights into their effects remain fragmented due to variations in theoretical frameworks, outcome measures, and game designs.

**Objectives**: This systematic review synthesizes experimental and intervention studies examining the effects of video games on prejudice and intercultural competence, aiming to identify patterns across game types and outcomes.

**Methods**: Studies were identified through database searches using predefined inclusion criteria. Eligible studies employed an experimental design with a control group and outcome measures targeting prejudice (e.g., attitudes, stereotypes, intergroup bias) or intercultural competence (e.g., empathy, knowledge, culturally appropriate behaviour). Twenty-four studies met these criteria and were organized into four analytical groups: (1) collaborative gameplay, (2) player–character interaction and embodiment, (3) comparisons of different games with bias-related content, and (4) comparisons of game- versus text-based formats.

**Results and Conclusions**: Across studies, video games showed potential to improve intercultural competence and prejudice-related outcomes. Effects were most consistent for collaborative and player–character interaction and embodiment game designs; by contrast, the impact of games with bias-related content depended on their moral framing, and comparisons of games with text suggested that interactive formats can have positive effects but demand careful experimental control. We identify strengths and limitations of the current body of research and discuss future research directions and practical implications.

*Keywords:* Video game, Prejudice reduction; Intercultural competence; Systematic review; Digital game-based learning

**Video Games, Prejudice, and Intercultural Competence: A Systematic Review**

Many societies around the world are diversifying. For example, by 2044, it is projected that over half of the U.S. population will identify as part of a minority group—that is, any group other than non-Hispanic White (US Census Bureau, 2019). Additionally, migration to Europe has risen in recent years, driven by social and military crises in regions such as the Middle East, Africa and Ukraine (e.g., (OECD, 2019, 2025b). As members from ethnic, racial and cultural minority groups often face oppression and marginalization and are at a higher risk of experiencing discrimination (OECD, 2025a; Sue et al., 2007), there is a need to develop methods and frameworks that can support their positive development and address systemic inequities (e.g., Hewstone, 2015; Leslie & Flynn, 2024; Salter et al., 2018). At the same time, tools are needed to help reduce prejudices, particularly among ethnic, racial, and cultural majority group members (e.g., Tropp et al., 2022), and foster intercultural competence (e.g., Deardorff, 2006; Schwarzenthal et al., 2020). Prejudice is understood as a negative attitude toward individuals based solely on their membership in a particular social group. It typically involves emotional responses, such as dislike or fear, cognitive components like stereotypes, and behavioural tendencies that can lead to discrimination (Allport, 1954; APA Dictionary of Psychology, 2023; Myers & Twenge, 2022). Intercultural competence has often been described a person's ability to interact effectively and appropriately with people from different cultural backgrounds (e.g., Chiu et al., 2013; Deardorff, 2004; Deardorff, 2006; Leung et al., 2014). Intercultural competence has been characterized by a multitude of attributes across the literature (Deardorff, 2004), including cognitive components (e.g., knowledge about norms and behaviours in different cultures), motivational components (e.g., enjoying interacting with people from diverse backgrounds), behavioural components (e.g., using appropriate behaviour in culturally diverse settings), and metacognitive components (e.g., recognizing that culture can influence one's own and others' behaviour and ways of thinking) (Ang et al., 2007; Van Dyne et al., 2012; see also (Matsumoto & Hwang, 2013). Across frameworks, additional commonly referenced aspects include empathy, respect, openness, and adaptability (see Deardorff, 2004).

Digital technologies, particularly video games, have been highlighted as effective means of engaging individuals from diverse backgrounds, offering immersive, interactive, and personalized settings (Arrigoni & Galani, 2019). If carefully designed and implemented, video games hold promises for reducing prejudice and fostering intercultural competences. There are several well-established theories that support this notion. First, the social-psychological intergroup contact theory suggests that positive interactions between members of different

ethnic, racial, and cultural groups can reduce prejudice, foster mutual understanding, and improve perceptions of outgroup members (e.g., Allport, 1954; Paolini et al., 2024; Tropp et al., 2022). While traditional contact occurs face-to-face, research has demonstrated that virtual interactions can also produce such positive effects (e.g., Amichai-Hamburger & McKenna, 2006; Costa et al., 2024; Dovidio et al., 2017). Also, video games can create shared digital spaces where players engage in cooperative gameplay, encouraging collaboration and mutual goal achievement, two key conditions for fostering positive intergroup relations (Allport, 1954; Paolini et al., 2024). Additionally, by removing social barriers, virtual contact in games may reduce intergroup anxiety, making cross-cultural interactions smoother and more comfortable (Tropp et al., 2022). This is particularly valuable in contexts where real-world contact is limited or marked by tension. For example, Benatov et al. (2021) demonstrated that an intervention using cooperative Minecraft gameplay significantly reduced intergroup bias and increased willingness for intergroup contact among Jewish-Israeli and Palestinian-Israeli children.

Beyond intergroup contact and prejudice reduction, video games may also enhance intercultural competence. For example, theory and research on multiculturalism (e.g., Bardach et al., 2024; Rosenthal & Levy, 2010; Schwarzenthal et al., 2020) propose that environments with a focus on learning about different cultures and perspectives promote positive intercultural outcomes, including intercultural competence. In this context, video games provide a unique platform for the seamless integration of multiculturalism elements. In video games, players often assume the roles of characters from diverse backgrounds, allowing them to experience different cultural narratives in an interactive and immersive way (Benatov et al., 2021). For instance, avatar embodiment, where players control or identify with an outgroup character, been discussed as a promising way to foster empathy and perspective-taking (Behm-Morawitz, Pennell, et al., 2016). Video games may also provide a safe environment for experimenting with identity and cultural norms by allowing players to interact, cooperate, or take on roles associated with outgroup members without real-world social consequences (Gjicali et al., 2020). Video games make intercultural learning more engaging by allowing players to experience different cultures through gameplay and storytelling. By interacting with diverse social environments and values, players gain firsthand experience that is often difficult to achieve in a formal educational setting. In this way, video games hold the potential to create learning opportunities that mirror real-world cross-cultural interactions, helping players navigate intercultural situations more effectively (Johnson, 2009). As players engage in problem-solving, dialogue, and narrative choices, they may develop practical intercultural skills that potentially extend

beyond the game world into real-life interactions.

However, despite the promise of video games, several challenges complicate a clear understanding of their impact. Specifically, the field is highly heterogeneous; for example, some studies on video games and prejudice reduction focus on format comparisons (e.g., video game versus text, H.-P. Chen et al., 2010), others examine the effects of avatar embodiment and character interaction (Behm-Morawitz, Pennell, et al., 2016), while still others compare different types of video games (Saleem & Anderson, 2013). This diversity highlights the need for a systematic review that synthesizes these lines of experimental research and provides in-depth insights into their effects on prejudice reduction and on intercultural competence outcomes. Relatedly, the specific games employed in these studies may play a central role in shaping outcomes. Thus, a detailed account of their design features—and discussion of how these features may be linked to observed effects—is essential for advancing the field. Also, to inform future research and practice, a critical reflection on the strengths and limitations of the current body of research seems warranted. Yet, to the best of our knowledge, such an overview has not been undertaken, perhaps due to the scattered nature of video game research in this area. The present systematic review sets out to address this gap.

The work most closely related to our systematic review is the meta-analysis by Costa et al. (2024), which synthesized research on diverse forms of digital intergroup contact (e.g., computer-mediated communication, non-playable character [NPC][1] exposure, avatar embodiment). While Costa and colleagues demonstrated that digital contact can reduce prejudice, their synthesis was broad in scope and focused primarily on prejudice outcomes, combining experimental and correlational designs. In contrast, the present review narrows the lens to experimental and intervention studies using video games. Moreover, we do not only consider prejudice reduction but also the promotion of intercultural competences, and pay close attention to the design features of the games themselves and their role in shaping outcomes.

The present systematic review addressed the following preregistered research questions (RQs): What are the characteristics of video games used in experimental and intervention research on prejudice reduction and intercultural competence (RQ1)? We distinguish between the broader categories of self-developed versus commercial games, and assign games to different genres such as simulation and narrative games, first-person shooters, third-person

[1] Non-playable characters (NPCs) are characters in a video game that are not controlled by the player but instead operate based on the game's programming, often serving as opponents, allies, or background figures within the game world.

action-adventure, sandbox, sports, strategy, and arcade-style games.

What are the effects of playing video games on prejudice and intercultural competence outcomes ((RQ2)? Here, we synthesize and discuss the effects reported in the studies included in our review. Given the highly heterogenous body of research, we group studies with similar features together which facilitates integration and the interpretation of results. Specifically, we distinguish between the following categories of studies: studies focusing on (a) collaborative gameplay, (b) avatar and character interaction, (c) different games with bias-related content, and (d) comparisons of video games versus text-based formats.

Lastly, we ask: What are the strengths and limitations of the current body of research (RQ3)? What are the most pressing and promising directions for future research? What potential practical implications can be derived from the current state of experimental and intervention research on video games, prejudice, and intercultural competences for different stakeholders (e.g., video game designers, educational practitioners) (RQ4)?

## Methods

### Literature Search

A systematic literature search in Web of Science (titles, abstracts, keywords) and PsycINFO (abstracts) was performed on 07/10/2024 using search terms related to video games, and intercultural competence and prejudice (see Table 1).

**Table 1**

*Search Strings and Data Base Search Results*

| Category | Search String (keywords) | Hits (before duplicates) | Hits (after duplicates) |
|---|---|---|---|
| Video Games (**first search string**) | (videogam OR video-gam* OR "video gam*" OR “gaming” OR “serious gam*” OR “digital gaming app*”) | - | - |
| Intercultural Competence & Prejudice (**second search string**) | (“intercultural competenc” OR “intercultural skill*” OR “intercultural learn*” OR “intercultural awareness“ OR “intercultural sensitivity” OR “intercultural intelligence“ OR “intercultural knowledge“ OR “multicultural competenc*” OR “multicultural skill*” OR “multicultural learn*” OR “multicultural awareness“ OR “multicultural sensitivity” OR “multicultural intelligence“ OR “multicultural knowledge“ OR “cross-cultural competenc*” OR “cross-cultural | - | - |

| Category | Search String (keywords) | Hits (before duplicates) | Hits (after duplicates) |
|---|---|---|---|
| | skill*" OR "cross-cultural learn*" OR "cross-cultural awareness" OR "cross-cultural sensitivity" OR "cross-cultural intelligence" OR "cross-cultural knowledge" OR "cultural competenc*" OR "cultural skill*" OR "cultural learn*" OR "cultural awareness" OR "cultural sensitivity" OR "cultural intelligence" OR "cultural knowledge" OR "sociocultural empathy" OR "ethnocultural empathy" OR "global competenc*" OR "intercultural communication skill*" OR "intercultural communication competenc*" OR "multicultural communication skill*" OR "multicultural communication competenc*" OR "cross-cultural communication skill*" OR "cross-cultural communication competenc*" OR "cultural communication skill*" OR "cultural communication competenc*" OR prejudice* OR stereotyp* OR "outgroup evaluation*" OR "outgroup attitude*" OR "intergroup attitude*" OR "intergroup bias*") | | |
| Combined (**AND**) | Both strings together | 934 (661 Web of Science; 273 PsycINFO) | 716 |

The two search strings were combined using the Boolean operator AND (Table 1). The first string covered video games, the second focused on intercultural competence and prejudice. The combined search yielded 934 records (see Table 1). After removing duplicates, 716 unique records remained.

**Study Inclusion and Expanded Search**

The titles and abstracts of all articles were independently screened by the first author and a trained research assistant, applying a set of inclusion and exclusion criteria. We included studies that fulfilled the following criteria: (a) The study reported an experiment or intervention (with an intervention and control group); (b) The study focused on effects of video games on prejudice (with respect to ethnicity, race, cultural group memberships) and/or intercultural competences. We excluded studies if they fulfilled one or more of the following criteria: (a) The abstract, full text, or secondary sources reporting the results of the study were not available;

(b)The study was not published in a peer-reviewed journal or book. Thus, for example, we excluded preprints, dissertations, and conference proceedings. The two independent screeners showed a good inter-rater agreement (98.39%), and all discrepancies regarding decisions to include versus exclude a specific article were discussed until consensus was reached. A total of 24 articles from the databases were included. The database search was then expanded by screening the references of previously published reviews and meta-analyses on related topics and expert recommendations. This expanded search did not yield any additional articles. The first author, the second author, and the fourth author extracted data from the full texts and coded information on the sample, the video game, the study design, the outcome variables, and the analytical approach and main findings. Any discrepancies in their codings were discussed until consensus was reached. In a final step, the fifth author checked all codings.

A total of 24 studies were included in the review. These studies were published between 2002 and 2025, and comprised between 65 and 433 participants, with a total of 4,392 participants. Most studies were conducted in the United States (16 studies), followed by Canada (2 studies). The remaining studies were carried out in Israel–Palestine, Germany, Taiwan, Singapore, Belgium, and Spain (one study each). Across all 24 studies, 69.6% of participants were university students (n = 3,058), 5.0% were sixth-grade students (n = 219), and 25.4% were members of the general population (n = 1,115). Females represented 57.74% of the total sample, males accounted for 41.09%, and 0.57% did not identify with either category. Across the 4,392 participants, reporting practices varied substantially across studies and included a mix of racial, ethnic, national, and religious identity categories. With respect to race and ethnicity, 46.76% identified as White, 4.15% as African American/Black, 3.65% as Asian, 2.34% as Hispanic/Latino, and 3.04% as another race/ethnicity. In addition, some studies reported religious or national identity, including Jewish (1.01%), Palestinian (1.01%), Taiwanese (2.96%), and Singaporean (6.83%). For 28.54% of participants, no information on race, ethnicity, nationality, or religion was reported.

**Transparency and Openness**

Our systematic review, including all research questions, literature search strategy, and inclusion and exclusion criteria, was preregistered on the Open Science Framework (OSF; anonymized link for peer review: https://osf.io/ebujw/overview?view_only=a48d45a13c174c4aa4f6acc8a8c3fe4f). There was one deviation from the preregistered inclusion criteria. We had originally specified that eligible studies must include both intervention and control groups with pre- and post-test measures.

During the review process, however, we decided to also include studies with intervention and control groups that assessed outcomes only at post-test. This decision was motivated by the observation that a substantial number of experimental studies employed very brief gameplay exposures and therefore did not include pre-test assessments, despite meeting all other core methodological standards. Excluding these studies would have resulted in the systematic omission of a sizeable and theoretically relevant segment of the literature.

## Results and Discussion

In the following, we present the results for each research question, alongside their interpretation within an integrated discussion.

### Characteristics of the Included Video Games (RQ1)

The studies included in this systematic review employed a diverse range of video games, which can be broadly divided into two categories: self-developed ($n$ = 7) and commercial ($n$ = 21). While self-developed games allow researchers precise control over content and variables, commercial games span a wide range of genres and spare researchers from time-intensive development of custom games.

Self-developed games are tailored to support experimental designs that examine specific research goals. These games often simulate real-life scenarios or fictional conflicts designed to evoke emotional and cognitive responses. For instance, *FairPlay* (Gutierrez et al., 2014) is a point-and-click narrative that places players in the role of a Black student navigating subtle racial bias in a university setting, while *Vox Populi* (Gjicali et al., 2020) creates a fictional political world with interactive cultural systems. Other examples include *Hire Up* (V. Chen et al., 2025), a management game featuring moral decision-making around NPCs of varied ethnic and national backgrounds, and *FORmosaHope* (H.-P. Chen et al., 2010), which uses exploration to foster awareness of Taiwanese culture and identity. Other games that were purpose-built for specific research projects included a fast-paced video game focusing on “shoot/don’t shoot” decisions developed by Correll et al. (2002, 2007), or a s*elf-developed 2D puzzle game* by Jarrell et al. (2021) designed to manipulate avatar features like race or gender.

Overall, even though self-developed video games offer adaptability to specific research questions, their development demands substantial time and resources for researchers. It is thus not surprising that the majority of studies included in this systematic review relied on commercially available titles, which are accessible, fully developed, and ready-to-use products. We next provide an overview of the main categories of commercial video games and the

specific video games examined in the studies included in this review.

Commercial games showed a wide variety of genres ranging from simulation and narrative games to first-person shooters, third-person action-adventure, sandbox, sports, strategy, and arcade-style games. Simulation and narrative games immerse players in bureaucratic, social, or ethical roles, encouraging moral reflection or identity-based decisions. For example, *PeaceMaker* (ImpactGames, 2007), a diplomacy simulation, has been used in research where participants take the role of either the Israeli Prime Minister or the Palestinian President (Alhabash & Wise, 2012, 2014). Similarly, *Papers, Please* (3909 LLC, 2013), a border control simulator, and *The Westport Independent* (Coffee Stain Studios, 2016), a censorship simulation about managing news reporting under government pressure, have been used to investigate how morally complex bureaucratic tasks influence empathy and policy attitudes (Peña et al., 2018; Peña & Hernández Pérez, 2020). Other studies examined commercial social simulation environments, such as *The Sims Social* (Playfish, 2011) which allowed researchers to explore how avatar race and social interactions shaped intergroup attitudes (Behm-Morawitz, Pennell, et al., 2016). *Second Life* (Linden Lab, 2003), a large-scale virtual environment where players design and embody avatars in flexible interactions without fixed scripts, was used to study how avatar identity and social context influence intergroup perceptions (Vang & Fox, 2014). Other narrative-driven titles include *Bury Me, My Love* (The Pixel Hunt, 2019), a smartphone-based interactive story where players guide a Syrian refugee through text messages and branching choices. This game has been used to examine empathy, intergroup attitudes, and willingness to help refugees (Lee & Chen, 2023).

First-person shooters (FPS) are typically fast-paced, combat-oriented games that immerse players in violent or high-stakes scenarios (Despain, 2017; Qaffas, 2020). Commonly used titles in this research include *Call of Duty: Black Ops* (Treyarch, 2010), a military shooter with team-based and competitive gameplay modes. *Call of Duty* has, for example, been used to study the effects of cooperative versus competitive play on intergroup attitudes (Adachi et al., 2015, 2016). *Counter-Strike* (Valve, 2000), a tactical FPS where players take on the roles of terrorists or counter-terrorists, was examined here for its potential to reinforce stereotypes of Arab outgroups (Saleem & Anderson, 2013). *Unreal Tournament 2004* (Epic Games, 1999), an arena-style shooter with customizable avatars and high-speed gameplay was employed to investigate how avatar race and embodiment influence aggression and social evaluation (Eastin et al., 2009).

Third-person action-adventure games allow players to see and control their avatar in the

game world, which can enhance embodiment and character-based decision-making. Examples include *Saints Row 2* (Volition, 2008), an open-world crime action game with customizable protagonists, that has been used by several studies included in our review to investigate how avatar race and violent gameplay influence intergroup attitudes (Behm-Morawitz, Hoffswell, et al., 2016; Yang et al., 2014). *The Elder Scrolls V: Skyrim* (Bethesda Game Studios, 2011), a fantasy role-playing game with extensive character customization that, in the case of research included in our review (Breves, 2020) was examined in a virtual reality (VR) setting to study the role of NPC interactions in shaping prejudice.

Sandbox games provide flexible environments without fixed goals for exploration and collaborative creation (Despain, 2017; Qaffas, 2020). A popular example is *Minecraft* (Mojang Studios, 2009), a block-based building and survival game that supports creative, co-operative problem-solving; it has been used to study intergroup collaboration between Jewish-Israeli and Palestinian-Israeli children (Benatov et al., 2021)

Sports games simulate real-world sports and team dynamics. For instance, *NBA 2K14* (Visual Concepts, 2013), a basketball simulator with both team-based and competitive modes, was applied to test cooperative versus solo play on intergroup attitudes (Adachi et al., 2016). *Fight Night Round 4* (EA Sports, 2009), a boxing game allowing avatar race customisation and one-on-one competition, was used to explore effects of avatar race on bias (Ash, 2016). Similarly, *WWE Smackdown vs. Raw 2010* (Yuke's Tose, 2009), a professional wrestling simulation featuring customizable characters and competitive matches, was used to examine how avatar ethnicity and violent play influence implicit and explicit racial attitudes (Yang et al., 2014).

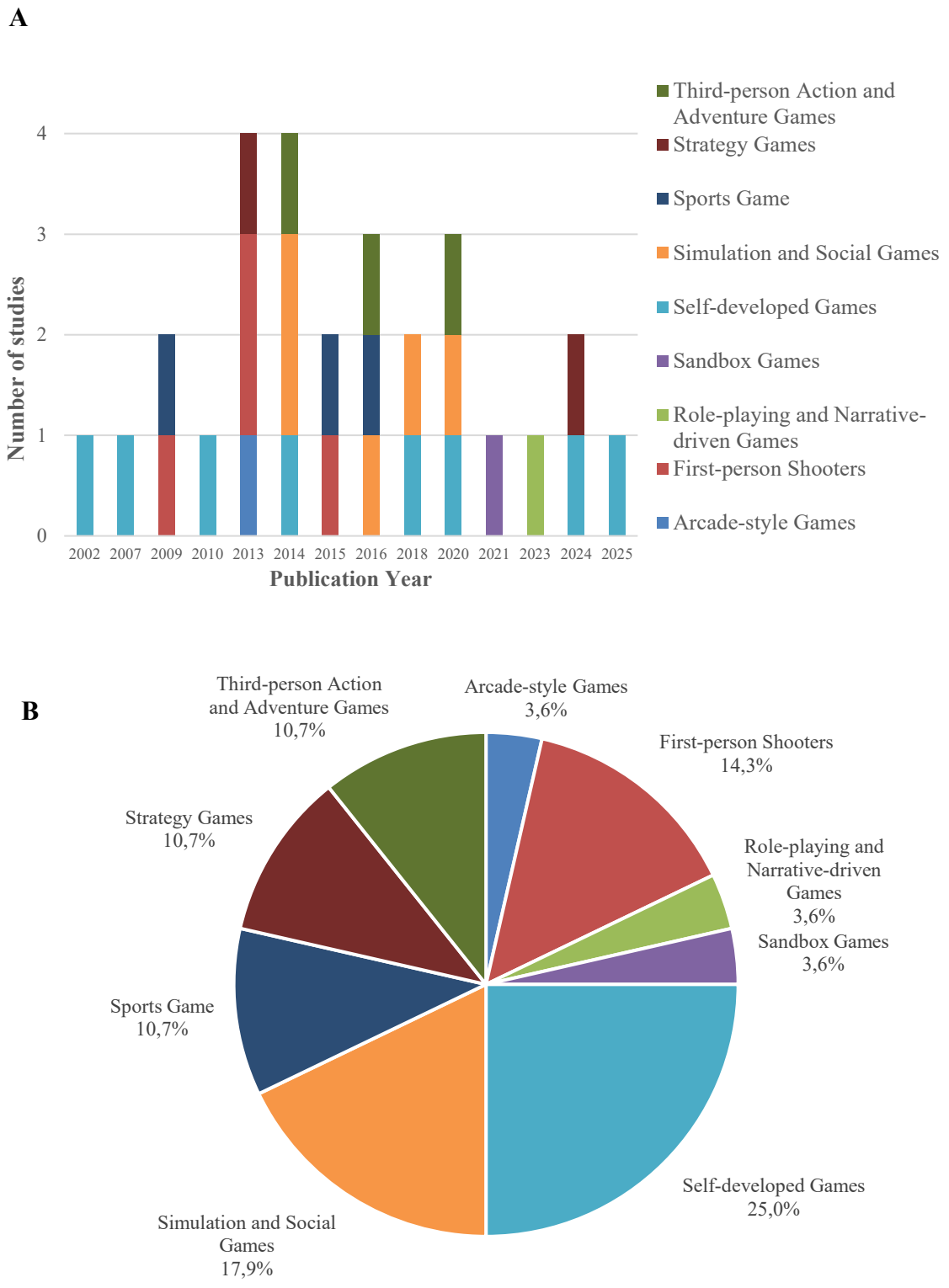
Strategy games engage players in planning, decision-making, and managing complex systems (Despain, 2017; Qaffas, 2020). An example is *Arabian Lords* (BreakAway Games, 2007), a resource-management game set in Middle Eastern cities between the 7th and 13th centuries, with a strong focus on Islamic history. This game has been employed to examine bias when Arab characters are included (Saleem & Anderson, 2013). Another example is the pair of *unnamed commercial Massive Online Battle Arena* games used in Pech and Caspar's (2024) study. The test game included bias-relevant narrative and gameplay structures intended to target prejudice, while the control game provided the same competitive framework but without any intergroup or bias-related elements. Massive Online Battle Arena games are generally considered a subgenre of strategy games (Mora-Cantallops & Sicilia, 2018).

Arcade-style games are typically quick to learn and play, emphasizing immediate feedback over complex narrative or systems (Despain, 2017; Qaffas, 2020). For example, *3D Ultra Mini Golf Adventures* (Wanako Games, 2006) is a casual mini-golf title built around short sessions and simple mechanics and has been used to study bias-related outcomes (Saleem & Anderson, 2013).

A complete overview of all games, their categories, and their research applications is provided in Appendices A-D and further elaborated in the discussion of RQ2. Figure 1 (Panel A) provides a descriptive overview of the video games included in our review across publication years. Commercial games predominate, while self-developed games appear consistently across years. We further note that in the studies included in our review, first-person shooter and sports games were used up to 2016, whereas simulation and social games as well as third-person action and adventure games were most often used between 2014 and 2020. Figure 1 (Panel B) illustrates the overall distribution of game genres, with simulation and narrative games being most common. At a descriptive level, genres also align with typical research applications: simulation/narrative titles are most often used for role-taking and intercultural scenarios; first-person shooters are typically chosen for high-time-pressure, stereotype-relevant decision contexts; sandbox and sports titles are used to study cooperation and team dynamics; and virtual or VR environments for manipulating avatar identity and player embodiment.

**Figure 1**

*Video games used in prejudice reduction and intercultural competence studies: Publications by year and genre (Panel A) and overall distribution of game genres (Panel B)*



**Effects of Video Games on Prejudice and Intercultural Competence Outcomes (RQ2)**

We summarize the effects of video games on prejudice and selected components of intercultural competence, across four groups of studies: collaborative gameplay (n = 3), avatar and character interaction (n = 15), different games with bias-related content (n = 3), and comparisons of video games versus text-based formats (n = 3). Across this literature, prejudice-related outcomes were frequently studied, whereas intercultural competence outcomes were rarely investigated, with some notable exceptions, especially in studies focusing on format comparisons. In our discussion below, we focus on the main effects of interest but report details on the samples, measures (including additional measures), and analyses in the Appendix.

**Collaborative gameplay.** First, three studies focused on collaborative gameplay, examining how cooperation with outgroup members shapes prejudice and intergroup dynamics. Appendix A reports a detailed overview of these studies' samples, video games used, design, outcomes, and findings. In two studies, Adachi et al. (2015, 2016) relied on samples of Canadian university students and investigated the effects of cooperative versus competitive gameplay on prejudice (attitudes towards different outgroups, including national outgroups [Americans] and university outgroups [other university]). The participants played *Call of Duty: Black Ops* (Adachi et al., 2015; Adachi et al., 2016, Study 1) and *NBA 2K14* (Adachi et al., 2016, Study 2). Adachi et al. (2015) showed that Canadian students' attitudes toward both Americans and towards students of the other university became more favourable after cooperating with an outgroup member, with large effect sizes. Adachi et al. (2016) replicated this pattern: in Study 1 (*Call of Duty: Black Ops*), intergroup cooperation (vs. solo or ingroup cooperation) led to reduced prejudice, measured as more favourable outgroup attitudes. In Study 2 (*Call of Duty: Black Ops* vs. *NBA 2K14*), cooperation again reduced prejudice, with additional analyses indicating that this effect was mediated by recategorization (but not empathy towards an outgroup member). Results did not differ between violent and non-violent games. Benatov et al. (2021) tested cooperative *Minecraft* play among Jewish-Israeli and Palestinian-Israeli 6th graders. Compared to control groups, intergroup teams showed reduced prejudice (e.g., lower negative stereotyping, less emotional prejudice) and greater willingness for contact, with effects remaining stable at six-month follow-up.

**Avatar and character interaction.** Second, 15 studies focused on avatar and character interaction as shown in Appendix B, exploring how avatar customisation, racial representation, and morally significant choices shape players' attitudes toward other social groups. Several studies used commercially available games with built-in options for avatar identity

manipulation. Alhabash and Wise (2012, 2014) examined the impact of playing *PeaceMaker*, where participants acted as either the Israeli Prime Minister or the Palestinian President. In both studies, explicit national attitudes (i.e., favourability and stereotypes toward Israelis and Palestinians) shifted depending on role assignment. In Alhabash and Wise (2012), playing the Palestinian President increased favourability toward Palestinians and decreased favourability toward Israelis, while playing the Israeli Prime Minister showed no significant change in favourability. In Alhabash and Wise (2014), playing the Palestinian President only led to more explicit stereotypes against Israelis, but not fewer stereotypes against Palestinians, and vice versa. Implicit attitudes and stereotype did not significantly change in either study.

Other research examined fighting and action titles such as *Fight Night Round 4* (Ash, 2016; Yang et al., 2014), *Saints Row 2* (Behm-Morawitz, Hoffswell, et al., 2016; Yang et al., 2014), and *WWE Smackdown vs. RAW 2010* (Yang et al., 2014), which were used to investigate how avatar race, embodiment, and violent gameplay influence intergroup attitudes. Ash (2016) tested *Fight Night Round 4* with a predominantly White student sample and found mostly null results, with avatar race not predicting stereotypes. Yang et al. (2014), however, reported that White students playing violently as a Black avatar in *Saints Row 2*, *WWE Smackdown vs. RAW*, and *Fight Night Round 4* expressed more negative explicit attitudes toward Black people and showed stronger implicit associations of Black people with weapons. Aggression also increased in these conditions, an effect that was partly explained by heightened implicit bias. Behm-Morawitz, Hoffswell, et al. (2016) found that White students assigned a Black avatar in *Saints Row 2* expressed fewer positive beliefs about African Americans than students in the control group, suggesting reinforced negative stereotyping. In contrast, Behm-Morawitz, Pennell, et al. (2016) used *The Sims Social* and showed that White students playing as a Black avatar reported more favourable beliefs about Black men and stronger support for pro-minority policies, especially when they liked their avatar. Taken together, these studies suggest that, among White participants, embodying Black avatars in violent contexts can reinforce negative stereotypes and prejudice, whereas embodying such avatars in social or everyday contexts may foster more positive intergroup beliefs and policy support. At the same time, some work (e.g., Ash, 2016) found largely null effects, underlining that outcomes are not uniform across studies.

Other studies used immersive or virtual environments to examine prejudice reduction. Breves (2020) tested *The Elder Scrolls V: Skyrim* in VR versus 2D and a non-game control with German students. They found that receiving a quest from a Black NPC in VR lowered explicit prejudice relative to control, with no effect on implicit prejudice. Vang and Fox (2014) used

*Second Life* with White U.S. undergraduates and found that partners represented by Black avatars were evaluated more positively than partners represented by White avatars; cooperation versus competition did not change this pattern. Jarrell et al. (2021) employed a *self-developed 2D puzzle game* with White male participants and assessed bias in a hiring task. Overall, female and Black applicants were rated as less competent than male and White applicants, and avatar race did not reduce this bias. However, higher embodied presence when playing a Black avatar was associated with more favourable evaluations of a Black female applicant. Taken together, immersive and virtual settings show selective benefits for prejudice-related outcomes—VR contact with a Black NPC improved explicit attitudes in one study, while effects in other contexts were modest or contingent on factors such as embodied presence.

Other studies examined self-developed or commercial games with different forms of outgroup representation. V. Chen et al. (2025) used *Hire Up*, a management game where players made moral or non-moral decisions about hiring "citizens" and "foreigners". While moral framing increased feelings of guilt, the presence of moral decisions did not influence post-game attitudes toward immigrants; in the condition in which participants interacted with both citizens and foreigners, they even reported less favourable attitudes toward immigrants. Correll et al. (2002, 2007) employed a *self-developed "shoot/don't shoot" game* and relatively consistently found a shooter bias across experiments. For example, white participants were quicker to shoot armed black targets than white targets and more likely to mistakenly shoot unarmed black targets than white targets. Eastin et al. (2009) used *Unreal Tournament 2004* with mirrored avatars to highlight race and found that identification was stronger when participants played as an avatar of their own race, with White participants identifying more with White avatars and Black participants more with Black avatars. Finally, Saleem and Anderson (2013) used a set of five games—namely, *Counter-Strike* with Arab terrorists, *Counter-Strike* with Russian terrorists, *Unreal Tournament 2004*, *Arabian Lords*, and *3D Ultra Mini Golf Adventures*—to manipulate the presence of Arab characters and terrorist themes. Their findings demonstrated that violent gameplay with Arab characters reinforced implicit and explicit anti-Arab attitudes, while non-violent or neutral games mitigated such effects. Overall, these findings suggest that in-game identity and interaction may influence how players perceive social groups, with violent outgroup portrayals potentially reinforcing prejudice, while non-violent or neutral contexts show more mixed or mitigating effects.

One notable exception within this group is the game-based social simulation *Vox Populi* (Gjicali et al., 2020). Unlike most studies focusing on prejudice, this study directly examined

components of intercultural competence. All participants played the same game about a fictional culture (Ustrad) and completed three quests to learn local customs, with interactions with NPCs recorded. They were randomly assigned to one of three conditions: Beliefs All, Beliefs at End, or No Beliefs. In the Beliefs All condition, participants rated their beliefs about Ustradian culture after each quest using a dropdown menu of statements; in the Beliefs at End condition, the same ratings were provided only after the final quest; and in the No Beliefs condition, no belief-rating prompts were given. The Beliefs All condition showed significantly more conversations with NPCs and received more correct feedback (indicating culturally appropriate actions) than the other two groups, and also achieved significantly higher cross-cultural concept knowledge than the No Beliefs condition. Overall, the results document that prompting players to reflect on their beliefs about the simulated culture increased social exploration, led to more culturally appropriate in-game actions, and improved players' conceptual understanding of key cross-cultural values (specifically norms related to social hierarchy and collectivism).

**Comparing games with bias-related content**. Third, three studies focused on different games with bias-related content (see Appendix C), investigating how varying gameplay contexts affect prejudice and related outcomes. Pech and Caspar (2024) employed a *custom-developed Massive Online Battle Arena* featuring two fictional factions to test whether long-duration gameplay influenced prosocial behaviour toward outgroup members and bias-related neural responses. They found that players of the test game showed more prosocial behaviour toward outgroup individuals and reduced bias-related neural responses compared to players of a *commercial Massive Online Battle Arena* without prejudice-related content (the control condition). Similarly, Peña et al. (2018) and Peña and Hernández Pérez (2020) contrasted *Papers, Please*—a morally complex immigration simulation—with *The Westport Independent*, a censorship-focused narrative game. Peña et al. (2018) tested U.S. undergraduates and found that intentions and self-efficacy to help immigrants decreased across gameplay sessions in both conditions, but remained higher among those who played *Papers, Please*, compared to those who played *The Westport Independent*. No reliable differences emerged for attitudes or subjective norms. Peña and Hernández Pérez (2020) replicated the study with Spanish undergraduates. Here, intentions to help immigrants and attitudes toward immigrants decreased over repeated sessions of *Papers, Please*; yet these outcomes remained more favourable than in the control group playing *The Westport Independent*. Measures of self-efficacy and subjective norms were excluded due to low reliability. Importantly, neither study included

empathy as an outcome, a noteworthy omission given that *Papers, Please* is often characterised as an "empathy simulator" (Isbister, 2016). Instead, Peña et al. (2018) assessed intentions, attitudes, subjective norms, and self-efficacy, while Peña and Hernández Pérez (2020) focused on intentions and attitudes. Taken together, these studies suggest that moral framing in immigration-focused gameplay may influence outcomes such as attitudes toward immigrants and intentions to help immigrants.

**Format comparisons.** Fourth, three studies compared video games with text-based formats (see Appendix D), isolating the impact of video game-based interactivity and immersion on empathy, bias awareness, and cultural understanding. Gutierrez et al. (2014) compared the *FairPlay* game, which places players in the shoes of a Black university student facing subtle racism, to a textual version of the same narrative. As hypothesized by the authors, they found that gameplayers with high empathy scores showed significantly lower implicit bias than gameplayers with lower empathy scores. However, it is also worth noting that, overall, the text condition had higher empathy scores than the game condition. Lee and Chen (2023) contrasted playing the mobile game *Bury Me, My Love* (either in the lab or at home) with reading news articles about Syrian refugees. Participants in both game conditions reported greater perceived closeness to refugees than those in the text condition. However, empathy and willingness to help refugees varied depending on context: participants playing at home reported lower empathy and willingness to help than those reading the news, while in-lab players did not differ significantly from the text condition. H.-P. Chen et al. (2010) tested *FORmosaHope* an educational game about Taiwanese culture, against a control group that received regular curriculum instruction. Students in the game condition reported significantly higher levels of cultural identity than the control group. These three studies show that interactive games may lead to stronger cultural identity, and, depending on the context and further variables, prejudice-related outcomes than non-interactive formats.

Taken together, the studies summarized in the four groups reveal that the effects of video games on prejudice and intercultural competences are varied and context-dependent. First, collaborative gameplay (Appendix A) reduced prejudice and improved intergroup attitudes when players cooperated with outgroup members. Second, studies focusing on avatar and character interaction (Appendix B) revealed that embodying avatars can either reinforce negative stereotypes and prejudice or foster more positive beliefs and policy support, depending on the gameplay context. Notably, outcomes varied according to whether the avatar experience reinforced or challenged prejudice, shaping perceptions of outgroups in negative or positive

ways. Third, research on bias-related games (Appendix C) suggests that moral framing and the presence of bias-related content can affect players' attitudes toward outgroups. Cooperative or reflective contexts tend to support attitudinal change and, in some cases, prosocial behaviour toward outgroup members, whereas violent contexts seem more likely to reinforce prejudice. Finally, in comparison to text-based formats, interactive games (Appendix D) may, under certain conditions, show an advantage in promoting emotional involvement and bias awareness. These patterns suggest that game mechanics, role structures, and interactivity may play a central role, themes that are further explored in the following sections.

**Strengths and Limitations of the Current Body of Research (RQ3)**

Over the past two decades, researchers have explored the potential of video games as tools for reducing prejudice and improving intercultural competence, though the number of studies, especially for intercultural competence outcomes, remains limited.

A key strength lies in the diversity of game types examined. Studies of collaborative gameplay (Appendix A) and investigations of avatar embodiment and perspective-taking (Appendix B) span both self-developed games tailored to specific research aims and commercially available titles across multiple genres. Further subsets of studies expand this diversity by examining games that vary in the extent and explicitness of bias-related content (Appendix C) and by contrasting interactive game formats with text-based versions of the same materials to isolate the unique contribution of interactivity and immersion (Appendix D). This diversity also highlights that different game designs may contribute to outcomes through distinct mechanisms: cooperative play often fosters prosocial behaviour, while role-taking and embodiment experiences can influence empathy and identity-related attitudes. Together, these approaches enable researchers to examine a wide range of processes relevant to prejudice reduction and changes in intercultural competence outcomes. The field also benefits from drawing on multiple theoretical perspectives, including intergroup contact theory (Benatov et al., 2021 in Appendix A) empathy development (Gutierrez et al., 2014; Lee & Chen, 2023 in Appendix D), and embodiment theory (Behm-Morawitz, Hoffswell, et al., 2016 in Appendix B), which provide useful frameworks for interpreting results.

However, the evidence base also shows notable limitations, many of which are visible across the four analytical groups. Numerous studies, especially those on avatar embodiment and perspective-taking (Appendix B) and studies on bias-related content (Appendix C), rely on small, convenient samples, often from university populations (e.g., Alhabash & Wise, 2012; Breves, 2020; Correll et al., 2007). Combined with the predominance of so-called minority

world contexts (i.e., world regions in which only the minority of the world's population resides, such as US and Europe, Alam, 2008; Khan et al., 2022; also referred to as WEIRD countries, Henrich et al., 2010), this restricts generalizability across age groups, socio-economic contexts, and cultural settings. Replication is also rare, with notable exceptions (e.g., Peña & Hernández Pérez, 2020), leaving many findings untested for robustness across samples and contexts.

Moreover, outcome measures focused primarily on prejudice and related attitudinal, affective, or implicit intergroup outcomes, rather than on intercultural competence (see e.g., Deardorff, 2004; Matsumoto & Hwang, 2013), which seems like a missed opportunity and an important area for intensified research efforts. Furthermore, we observed that several studies employed implicit measures such as Implicit Association Tests particularly in avatar and character interaction studies (e.g., Yang et al., 2014; Saleem & Anderson, 2013; Breves, 2020; Behm-Morawitz, Pennell, et al., 2016) and in one case in game–text comparison studies (Gutierrez et al., 2014). These tests are widely used to capture automatic associations underlying prejudice (even though they have also been criticized for issues of reliability and predictive validity; Oswald et al., 2013). Other studies relied on study-specific or custom-developed questionnaires that were not consistently validated across samples or studies. This inconsistency not only hinders cumulative evidence building but also makes it harder to identify systematic patterns. Establishing greater consistency in outcome measures will be essential for building a cumulative evidence base. In addition, the description of the games used was often limited, leaving unclear which specific elements (narrative, mechanics, or social interaction) drive the observed effects. Without clearer documentation, it remains difficult to disentangle which game features may be primarily responsible for changes in prejudice or in intercultural competence.

Another limitation concerns the range of age groups and study contexts. The vast majority of studies recruited university students or young adults, while only very few examined younger participants (e.g., Benatov et al., 2021 with sixth graders; H.-P. Chen et al., 2010 with elementary school students). This underrepresentation of children and adolescents limits insights into how video games may shape prejudice and intercultural competences at earlier developmental stages. Similarly, most studies were conducted in laboratory settings, with only a small number testing gameplay in more naturalistic, real-life contexts (e.g., Lee & Chen, 2023 with a take-home condition; Pech & Caspar, 2024 with at-home sessions). Moreover, with some exceptions (e.g., Benatov et al., 2021), longer-term intervention studies, especially those including follow-up assessments remain scarce, leaving the long-term effects of gameplay

largely unexamined. In addition, most experiments relied on brief, single play sessions (often ~12–30 minutes), whereas only a few studies implemented substantially longer or repeated exposure (e.g., 2.5–4 hours in Pech & Caspar, 2024). These differences raise questions about how long-term or more real-world exposure might affect outcomes, pointing to the need for systematic research on dosage effects and real-world play conditions. Addressing these limitations is critical for advancing the field, enabling more rigorous, generalizable, and actionable insights into how video games can reduce prejudice and influence intercultural competence outcomes, and how specific game mechanics may contribute to these effects.

**Future Research Directions and Practical Implications (RQ4)**

Findings across the four groups highlight specific priorities. Collaborative gameplay (Appendix A) consistently fostered the investigated outcomes, but only three studies were included into this category. Thus, intensified research efforts are needed, and we consider a stronger focus on interventions in real-life contexts outside of the lab (e.g., school classes) essential for advancing research in this area. Avatar and character interaction (Appendix B) showed promising but context-dependent effects, which reinforces the need for sound replications to help building a robust evidence base. Comparisons of games with bias-related content (Appendix C) demonstrated that narrative framing shapes attitudes, underscoring the need to isolate moral and contextual factors. Comparisons of video games with text-based formats (Appendix D) showed that outcomes often depend on study design. To attribute effects to the format itself, we propose that game and text conditions may need to be carefully matched on content, duration, and context (e.g., Viccari & Bardach, 2025). This ensures that differences between video game and text formats reflect interactivity rather than differences in content. Insights from such research could then be used to guide the intentional design of future games, ensuring that the most effective mechanics are embedded from the beginning.

In general, future research on video game effects should broaden participant diversity by actively recruiting from different cultural, socio-economic, and age groups (e.g., Lui et al., 2021; Priego-Ojeda et al., 2025) to ensure that findings on prejudice and intercultural competence outcomes are globally relevant. Replicating studies across diverse cultural contexts will be essential to determine whether the findings we observed here are applicable to other groups ad contexts.

From a methodological perspective, priorities include conducting more extensive intervention studies and investigate long-term effects of such interventions. Future studies should adopt standardized and validated outcome measures for prejudice and other intergroup

or intercultural outcomes, facilitating comparability across projects and enabling robust meta-analyses. Another key priority is to examine which game mechanics drive the observed effects. We envision studies that systematically vary and isolate elements such as the level of interactivity, the narrative framing, or whether play is cooperative or competitive, to identify which aspects have the strongest impact on reducing prejudice and shaping intercultural competence outcomes.

The findings of our review and the outlined research directions carry implications for practice. For game designers, evidence from studies focusing on avatar and character interaction and perspective-taking suggests that building games that encourage players to take the perspective of others, work together to solve problems, and explore diverse cultural settings can increase the likelihood of positive attitudinal change. Cultural representation should be handled with care, ensuring it is authentic and nuanced through consultation with relevant communities to avoid reinforcing stereotypes. Designers can also strengthen the impact of their games by including prompts or discussion elements that encourage players to reflect on both the benefits and limitations of their in-game experiences (Ke, 2016). These could take the form of short debrief panels, in-game rating systems, or brief questions with no predefined answers, allowing participants to express key words and impressions. Such features not only support players' understanding and critical engagement but also provide valuable feedback to inform future design.

For educational practitioners, the choice or adaptation of games should be closely aligned with clearly defined learning objectives related to prejudice reduction and intercultural understanding. Guided reflection, whether during or after gameplay, can facilitate deeper processing and encourage the application of in-game experiences to real-world contexts. This can be operationalized through classroom discussions, structured debrief sessions, or short reflective writing tasks. Attention should also be paid to the cultural relevance of game narratives, ensuring that they resonate with the specific learner groups and contexts in which they are used. When developing games specifically for educational purposes, designers should aim to balance pedagogical goals with the usability, engagement, and production standards common in commercial titles, learning from industry practices and systematically integrating user feedback. Commercial games, by contrast, can be easier to implement but often have fixed elements that cannot be altered. Despite these constraints, their wide reach makes them valuable tools for scalable interventions, especially if they are embedded within carefully structured educational programs.

For policymakers and funders, promoting sustained collaboration between researchers, educators, and developers is essential for producing high-quality games and related research (Bucchiarone et al., 2025). To reach this goal, it would be desirable to prioritize enabling conditions such as infrastructure for cross-sector partnerships, stable funding streams, and incentives for replication studies. Ideally, continuity should be encouraged, with funding focused on the refinement, expansion, and replication of existing projects rather than defaulting to entirely new initiatives. Investment should also emphasize long-term evaluation rather than one-off projects, which could help identifying sustainable and scalable ways to use video games for prejudice reduction and the promotion of intercultural competence outcomes. Pursuing these directions may offer clearer guidance to those seeking to apply video games for social and educational applications.

## Conclusions

This review examined whether and how video games can reduce prejudice and influence intercultural competences. Overall, the evidence indicates that games and specific features (e.g., cooperative play) can counteract prejudice and promote intercultural competence outcomes, but these effects are not universal. Outcomes depend strongly on design choices, the social and narrative context of play, and the way games are embedded within research or educational settings.

Across the body of work, two consistent mechanisms stand out: collaboration with outgroup members and perspective-taking through avatars or characters tend to foster more positive intergroup attitudes. By contrast, the impact of games with bias-related content depends heavily on their moral framing, and comparisons of games with text highlight that interactive formats can have benefits. These patterns underscore that the medium itself is not inherently beneficial or harmful; rather, it is the specific mechanics and contexts that determine whether games reduce or reinforce prejudice and provide or hamper meaningful intercultural learning experiences.

## References

References marked with an asterisk indicate studies included in the systematic review.

3909 LLC. (2013). *Papers, Please* [Computer software]. 3909 LLC. https://papersplea.se/

*Adachi, P., Hodson, G., Willoughby, T., Blank, C., & Ha, A. (2016). From Outgroups to Allied Forces: Effect of Intergroup Cooperation in Violent and Nonviolent Video Games on Boosting Favorable Outgroup Attitudes. *Journal of Experimental Psychology: General*, *145*(3), 259–265. APA PsycInfo. https://doi.org/10.1037/xge0000145

*Adachi, P., Hodson, G., Willoughby, T., & Zanette, S. (2015). Brothers and Sisters in Arms: Intergroup Cooperation in a Violent Shooter Game Can Reduce Intergroup Bias. *Psychology of Violence*, *5*(4), 455–462. APA PsycInfo. https://doi.org/10.1037/a0037407

ADL. (2004). *Attitudes Toward Jews, Israel and the Palestinian-Israeli Conflict in Ten European Countries* (New York, NY.) [Report]. Anti-Defamation League. https://archive.jpr.org.uk/object-2740

Agarwal, R., & Karahanna, E. (2000). Time Flies When You're Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage. *MIS Quarterly*, *24*(4), 665–694. https://doi.org/10.2307/3250951

Alam, S. (2008). Majority World: Challenging the West's Rhetoric of Democracy. *Amerasia Journal*, *34*(1), 88–98. https://doi.org/10.17953/amer.34.1.l3176027k4q614v5

*Alhabash, S., & Wise, K. (2012). PeaceMaker: Changing Students' Attitudes Toward Palestinians and Israelis Through Video Game Play. *International Journal of Communication*, *6*, 356–380.

*Alhabash, S., & Wise, K. (2014). Playing their game: Changing stereotypes of Palestinians and Israelis through videogame play. *New Media & Society*, *17*(8), 1358–1376. APA PsycInfo. https://doi.org/10.1177/1461444814525010

Allport, G. W. (1954). *The nature of prejudice* (pp. xviii, 537). Addison-Wesley.

Altemeyer, B. (1988). *Enemies of freedom: Understanding right-wing authoritarianism* (1st

ed). Jossey-Bass Publishers.

Amichai-Hamburger, Y., & McKenna, K. Y. A. (2006). The Contact Hypothesis Reconsidered: Interacting via the Internet. *Journal of Computer-Mediated Communication*, *11*(3), 825–843. https://doi.org/10.1111/j.1083-6101.2006.00037.x

Anderson, C. A. (1999). *Coding key for the word completion task*. Institution: Iowa State University, Department of Psychology.

Anderson, C. A., Carnagey, N. L., & Eubanks, J. (2003). Exposure to violent media: The effects of songs with violent lyrics on aggressive thoughts and feelings. *Journal of Personality and Social Psychology*, *84*(5), 960–971. https://doi.org/10.1037/0022-3514.84.5.960

Anderson, C. A., Carnagey, N. L., Flanagan, M., Benjamin Jr., A. J., Eubanks, J., & Valentine, J. C. (2004). Violent Video Games: Specific Effects of Violent Content on Aggressive Thoughts and Behavior. In *Advances in experimental social psychology, Vol. 36* (pp. 199–249). Elsevier Academic Press. https://doi.org/10.1016/S0065-2601(04)36004-1

Anderson, C. A., Deuser, W. E., & DeNeve, K. M. (1995). Hot Temperatures, Hostile Affect, Hostile Cognition, and Arousal: Tests of a General Model of Affective Aggression. *Personality and Social Psychology Bulletin*, *21*(5), 434–448. https://doi.org/10.1177/0146167295215002

Ang, S., Van Dyne, L., Koh, C., Ng, K. Y., Templer, K. J., Tay, C., & Chandrasekar, N. A. (2007). Cultural Intelligence: Its Measurement and Effects on Cultural Judgment and Decision Making, Cultural Adaptation and Task Performance. *Management and Organization Review*, *3*(3), 335–371. https://doi.org/10.1111/j.1740-8784.2007.00082.x

APA Dictionary of Psychology. (2023). *Career Counseling*. APA Dictionary of Psychology. https://dictionary.apa.org/career-counseling

Appiah, O. (2001). Ethnic identification on adolescents' evaluations of advertisements. *Journal of Advertising Research*, *41*(5), 7–22. https://doi.org/10.2501/JAR-41-5-7-22

Appiah, O. (2006). Rich Media, Poor Media: The Impact of Audio/Video vs. Text/Picture

Testimonial Ads on Browsers'Evaluations of Commercial Web Sites and Online Products. *Journal of Current Issues & Research in Advertising*, *28*(1), 73–86. https://doi.org/10.1080/10641734.2006.10505192

Appiah, O. (2007). The Effectiveness of "Typical-User" Testimonial Advertisements on Black and White Browsers' Evaluations of Products on Commercial Websites: Do They Really Work? *Journal of Advertising Research*, *47*(1), 14–27. https://doi.org/10.2501/S0021849907070031

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, *63*(4), 596–612. https://doi.org/10.1037/0022-3514.63.4.596

Arrigoni, G., & Galani, A. (2019). Digitally enhanced polyvocality and reflective spaces: Challenges in sustaining dialogue in museums through digital technologies. In *European Heritage, Dialogue and Digital Practices*. Routledge.

*Ash, E. (2016). Priming or Proteus Effect? Examining the Effects of Avatar Race on In-Game Behavior and Post-Play Aggressive Cognition and Affect in Video Games. *Games and Culture: A Journal of Interactive Media*, *11*(4), 422–440. APA PsycInfo. https://doi.org/10.1177/1555412014568870

Bachen, C. M., Hernández-Ramos, P. F., & Raphael, C. (2012). Simulating REAL LIVES: Promoting Global Empathy and Interest in Learning Through Simulation Games. *Simulation & Gaming*, *43*(4), 437–460. https://doi.org/10.1177/1046878111432108

Bardach, L., Röhl, S., Oczlon, S., Schumacher, A., Lüftenegger, M., Lavelle-Hill, R., Schwarzenthal, M., & Zitzmann, S. (2024). Cultural diversity climate in school: A meta-analytic review of its relationships with intergroup, academic, and socioemotional outcomes. *Psychological Bulletin*, *150*(12), 1397–1439. https://doi.org/10.1037/bul0000454

Bar-Tal, D., & Labin, D. (2001). The effect of a major event on stereotyping: Terrorist attacks in Israel and Israeli adolescents' perceptions of Palestinians, Jordanians and Arabs. *European Journal of Social Psychology*, *31*(3), 265–280. https://doi.org/10.1002/ejsp.43

Batson, C. D., Early, S., & Salvarani, G. (1997). Perspective Taking: Imagining How Another Feels Versus Imaging How You Would Feel. *Personality and Social Psychology Bulletin*, *23*(7), 751–758. https://doi.org/10.1177/0146167297237008

Behm-Morawitz, E. (2013). Mirrored selves: The influence of self-presence in a virtual world on health, appearance, and well-being. *Computers in Human Behavior*, *29*(1), 119–128. https://doi.org/10.1016/j.chb.2012.07.023

*Behm-Morawitz, E., Hoffswell, J., & Chen, S. (2016). The Virtual Threat Effect: A Test of Competing Explanations for the Effects of Racial Stereotyping in Video Games on Players' Cognitions. *Cyberpsychology, Behavior, and Social Networking*, *19*(5), 308–313. APA PsycInfo. https://doi.org/10.1089/cyber.2015.0461

*Behm-Morawitz, E., Pennell, H., & Speno, A. (2016). The effects of virtual racial embodiment in a gaming app on reducing prejudice. *COMMUNICATION MONOGRAPHS*, *83*(3), 396–418. APA PsycInfo. https://doi.org/10.1080/03637751.2015.1128556

*Benatov, J., Berger, R., & Tadmor, C. T. (2021). Gaming for peace: Virtual contact through cooperative video gaming increases children's intergroup tolerance in the context of the Israeli–Palestinian conflict. *Journal of Experimental Social Psychology*, *92*, 104065. https://doi.org/10.1016/j.jesp.2020.104065

Berger, R., Brenick, A., Lawrence, S. E., Coco, L., & Abu-Raiya, H. (2018). Comparing the effectiveness and durability of contact- and skills-based prejudice reduction approaches. *Journal of Applied Developmental Psychology*, *59*, 46–53. https://doi.org/10.1016/j.appdev.2018.04.002

Bethesda Game Studios. (2011). *The Elder Scrolls V: Skyrim* [Computer software]. Bethesda Softworks. https://store.steampowered.com/app/489830/The_Elder_Scrolls_V_Skyrim_Special_Edition/

BreakAway Games. (2007). *Arabian Lords* [Computer software]. Quirkat Interactive Entertainment. https://videogamegeek.com/videogame/89077/arabian-lords

*Breves, P. (2020). Reducing Outgroup Bias through Intergroup Contact with Non-Playable

Video Game Characters in VR. *Presence-Virtual and Augmented Reality*, *27*(3), 257–273. https://doi.org/10.1162/pres_a_00330

Bucchiarone, A., Rossi, V., & Wanick, V. (Eds.). (2025). *Engineering Educational Games for a Sustainable Society: Play, Learn and Transform*. Springer Nature Switzerland. https://doi.org/10.1007/978-3-032-02591-3

*Chen, H.-P., Lien, C.-J., Annetta, L., & Lu, Y.-L. (2010). The Influence of an Educational Computer Game on Children's Cultural Identities. *Educational Technology & Society*, *13*(1), 94–105. eric.

*Chen, V., Koek, W., & Bowman, N. (2025). The effect of in-game moral choices and NPCs' identities on players' intergroup attitudes. *Entertainment Computing*, *52*. https://doi.org/10.1016/j.entcom.2024.100766

Chen, X.-P., Yao, X., & Kotha, S. (2009). Entrepreneur Passion And Preparedness In Business Plan Presentations: A Persuasion Analysis Of Venture Capitalists' Funding Decisions. *Academy of Management Journal*, *52*(1), 199–214. https://doi.org/10.5465/amj.2009.36462018

Chiu, C.-Y., Lonner, W. J., Matsumoto, D., & Ward, C. (2013). Cross-Cultural Competence: Theory, Research, and Application. *Journal of Cross-Cultural Psychology*, *44*(6), 843–848. https://doi.org/10.1177/0022022113493716

Coffee Stain Studios. (2016). *The Westport Independent* [Computer software]. Coffee Stain Studios. https://www.doublezeroonezero.com/westport.html

*Correll, J., Park, B., Judd, C., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, *83*(6), 1314–1329. https://doi.org/10.1037//0022-3514.83.6.1314

*Correll, J., Park, B., Judd, C., & Wittenbrink, B. (2007). The influence of stereotypes on decisions to shoot. *European Journal of Social Psychology*, *37*(6), 1102–1117. APA PsycInfo. https://doi.org/10.1002/ejsp.450

Costa, L. P., Bierwiaczonek, K., & Bianchi, M. (2024). Does Digital Intergroup Contact Reduce

Prejudice? A Meta-Analysis. *Cyberpsychology, Behavior, and Social Networking*, *27*(7), 440–451. https://doi.org/10.1089/cyber.2023.0591

Deardorff, D. K. (2004). *The Identification and Assessment of Intercultural Competence as a Student Outcome of Internationalization at Institutions of Higher Education in the United States*. http://www.lib.ncsu.edu/resolver/1840.16/5733

Deardorff, D. K. (2006). Identification and Assessment of Intercultural Competence as a Student Outcome of Internationalization. *Journal of Studies in International Education*, *10*(3), 241–266. https://doi.org/10.1177/1028315306287002

Despain, W. (Ed.). (2017). *Writing for Video Game Genres: From FPS to RPG*. Taylor & Francis Ltd.

Dovidio, J. F., Love, A., Schellhaas, F. M. H., & Hewstone, M. (2017). Reducing intergroup bias through intergroup contact: Twenty years of progress and future directions. *Group Processes & Intergroup Relations*, *20*(5), 606–620. https://doi.org/10.1177/1368430217712052

Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, *23*(3), 316–326. https://doi.org/10.1177/0146167297233009

EA Sports. (2009). *Fight Night Round 4* [Computer software]. Electronic Arts. https://en.wikipedia.org/w/index.php?title=Fight_Night_Round_4&oldid=1288333786

Eastin, M. (2006). Video Game Violence and the Female Game Player: Self- and Opponent Gender Effects on Presence and Aggressive Thoughts. *Human Communication Research*, *32*(3), 351–372. https://doi.org/10.1111/j.1468-2958.2006.00279.x

*Eastin, M., Appiah, O., & Cicchirllo, V. (2009). Identification and the Influence of Cultural Stereotyping on Postvideogame Play Hostility. *Human Communication Research*, *35*(3), 337-U10. APA PsycInfo. https://doi.org/10.1111/j.1468-2958.2009.01354.x

Epic Games. (1999). *Unreal Tournament 2004* [Computer software]. GT Interactive. https://en.wikipedia.org/w/index.php?title=Unreal_Tournament&oldid=1307302350

Gaertner, S. L., & Dovidio, J. F. (2013). *Reducing intergroup bias: The common ingroup identity model*. Psychology.

García, M. C. G., Navas, M. S., Guirado, M. I. C., & Alonso, F. M. (2003). Immigration and prejudice: Attitudes of sample of adolescents from Almeria [in Spanish]. *Acción Psicológica*, *2*(2), 137–147. https://doi.org/10.5944/ap.2.2.527

*Gjicali, K., Finn, B. M., & Hebert, D. (2020). Effects of belief generation on social exploration, culturally-appropriate actions, and cross-cultural concept learning in a game-based social simulation. *Computers & Education*, *156*. https://doi.org/10.1016/j.compedu.2020.103959

Graham, J., Haidt, J., & Nosek, B. (2008). *The Moral Foundations Questionnaire*. Moral Foundations. https://moralfoundations.org/questionnaires/

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. https://doi.org/10.1037/0022-3514.85.2.197

Greitemeyer, T. (2013). Playing Violent Video Games Increases Intergroup Bias. *Personality and Social Psychology Bulletin*, *40*(1), 70–78. APA PsycInfo. https://doi.org/10.1177/0146167213505872

*Gutierrez, B., Kaatz, A., Chu, S., Ramirez, D., Samson-Samuel, C., & Carnes, M. (2014). "Fair Play'': A Videogame Designed to Address Implicit Race Bias Through Active Perspective Taking. *Games for Health Journal*, *3*(6), 371–378. https://doi.org/10.1089/g4h.2013.0071

Hartmann, T., Toz, E., & Brandon, M. (2010). Just a Game? Unjustified Virtual Violence Produces Guilt in Empathetic Players. *Media Psychology*, *13*(4), 339–363. https://doi.org/10.1080/15213269.2010.524912

Hartmann, T., Wirth, W., Schramm, H., Klimmt, C., Vorderer, P., Gysbers, A., Böcking, S., Ravaja, N., Laarni, J., Saari, T. A., Gouveia, F. R., & Sacau, A. M. (2016). The spatial presence experience scale (SPES): A short self-report measure for diverse media settings. *Journal of Media Psychology*, *28*(1), 1–15. https://doi.org/10.1027/1864-1105/a000137

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Henry, P. J., & Sears, D. O. (2002). The Symbolic Racism 2000 Scale. *Political Psychology*, *23*(2), 253–283. https://doi.org/10.1111/0162-895X.00281

Hewstone, M. (2015). Consequences of Diversity for Social Cohesion and Prejudice: The Missing Dimension of Intergroup Contact. *Journal of Social Issues*, *71*(2), 417–438. https://doi.org/10.1111/josi.12120

Hodson, G., Dovidio, J. F., & Esses, V. M. (2003). Ingroup identification as a moderator of positive–negative asymmetry in social discrimination. *European Journal of Social Psychology*, *33*(2), 215–233. https://doi.org/10.1002/ejsp.141

ImpactGames. (2007). *PeaceMaker* [Computer software]. ImpactGames. https://www.peacemakergame.com/

Isbister, K. (2016). Her Story and Papers, Please are changing gaming forever. *Wired*. https://www.wired.com/story/games-power-captivate-engage/

*Jarrell, M. A., Anaraky, R. G., Knijnenburg, B. P., & Ash, E. M. (2021). *Using Intersectional Representation & Embodied Identification in Standard Video Game Play to Reduce Societal Biases* (WOS:000758168001061). CHI ’21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3411764.3445161

Johnson, W. L. (2009). Developing intercultural competence through videogames. *Proceedings of the 2009 International Workshop on Intercultural Collaboration*, 99–100. https://doi.org/10.1145/1499224.1499226

Kaminsky, M., & Bar-Tal, D. (1996). Stereotypic perceptions of different Israeli-Arab labels as a function of age and religiosity. *Studies in Education*, *1*, 121–157.

Karmarkar, U. R., & Tormala, Z. L. (2010). Believe Me, I Have No Idea What I'm Talking About: The Effects of Source Certainty on Consumer Involvement and Persuasion. *Journal of Consumer Research*, *36*(6), 1033–1049. https://doi.org/10.1086/648381

Kawakami, K., & Dovidio, J. F. (2001). The reliability of implicit stereotyping. *Personality and Social Psychology Bulletin*, *27*(2), 212–225. https://doi.org/10.1177/0146167201272007

Ke, F. (2016). Designing and integrating purposeful learning in game play: A systematic review. *Educational Technology Research and Development*, *64*(2), 219–244. https://doi.org/10.1007/s11423-015-9418-1

Khan, T., Abimbola, S., Kyobutungi, C., & Pai, M. (2022). How we classify countries and people—And why it matters. *BMJ Global Health*, *7*(6). https://doi.org/10.1136/bmjgh-2022-009704

Lamm, C., Batson, C. D., & Decety, J. (2007). The neural substrate of human empathy: Effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience*, *19*(1), 42–58. https://doi.org/10.1162/jocn.2007.19.1.42

*Lee, Y., & Chen, M. (2023). Intergroup Contact With a Virtual Refugee: Reducing Prejudice Through a Cooperative Game. *Psychology of Popular Media*, *12*(4), 383–392. APA PsycInfo. https://doi.org/10.1037/ppm0000433

Leslie, L. M., & Flynn, E. (2024). Diversity Ideologies, Beliefs, and Climates: A Review, Integration, and Set of Recommendations. *Journal of Management*, *50*(3), 849–876. https://doi.org/10.1177/01492063221086238

Leung, K., Ang, S., & Tan, M. L. (2014). Intercultural Competence. *Annual Review of Organizational Psychology and Organizational Behavior*, *1*(Volume 1, 2014), 489–519. https://doi.org/10.1146/annurev-orgpsych-031413-091229

Lieberman, J. D., Solomon, S., Greenberg, J., & McGregor, H. A. (1999). A hot new way to

measure aggression: Hot sauce allocation. *Aggressive Behavior*, *25*(5), 331–348. https://doi.org/10.1002/(SICI)1098-2337(1999)25:5%253C331::AID-AB2%253E3.0.CO;2-1

Linden Lab. (2003). *Second Life* [Computer software]. Linden Lab. https://secondlife.com/

Lui, K., Yip, K., & Wong, A. (2021). Gender differences in multitasking experience and performance. *The Quarterly Journal of Experimental Psychology*, *74*(2), 344–362. APA PsycInfo. https://doi.org/10.1177/1747021820960707

Mackie, D. M., Devos, T., & Smith, E. R. (2000). Intergroup emotions: Explaining offensive action tendencies in an intergroup context. *Journal of Personality and Social Psychology*, *79*(4), 602–616. https://doi.org/10.1037/0022-3514.79.4.602

Mayer, J. D. (2004). Christian Fundamentalists and Public Opinion Toward the Middle East: Israel's New Best Friends? *Social Science Quarterly*, *85*(3), 695–712.

McConahay, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology*, *72*(2), 262–274. https://doi.org/10.1037/0022-3514.72.2.262

McConahay, J. B., Hardee, B. B., & Batts, V. (1981). Has Racism Declined in America?: It Depends on Who is Asking and What is Asked. *Journal of Conflict Resolution*, *25*(4), 563–579. https://doi.org/10.1177/002200278102500401

McKirnan, D. J., Smith, C. E., & Hamayan, E. V. (1983). A sociolinguistic approach to the belief-similarity model of racial attitudes. *Journal of Experimental Social Psychology*, *19*(5), 434–447. https://doi.org/10.1016/0022-1031(83)90021-5

Mojang Studios. (2009). *Minecraft* [Computer software]. Mojang Studios. https://www.minecraft.net/de-de

Mora-Cantallops, M., & Sicilia, M.-Á. (2018). MOBA games: A literature review. *Entertainment Computing*, *26*, 128–138. https://doi.org/10.1016/j.entcom.2018.02.005

Myers, D. G., & Twenge, J. M. (2022). *Social psychology* (Fourteenth edition). McGraw Hill Education.

Nosek, B. A. (n.d.). *Implicit Association Test (IAT): Race (Black/White) IAT*. Project Implicit. http://projectimplicit.net/nosek/iat/

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G., & Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*, 36–88. https://doi.org/10.1080/10463280701489053

Novo-Corti, I., Muñoz-Cantero, J.-M., & Calvo-Porral, C. (2011). Analysis of young college students' attitudes toward physical handicaps: A theory of planned behavior view [in Spanish]. *RELIEVE - Revista Electrónica de Investigación y Evaluación Educativa*, *17*(2). https://doi.org/10.7203/relieve.17.2.4010

OECD. (2019). *PISA 2018 Results (Volume II): Where All Students Can Succeed*. Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/education/pisa-2018-results-volume-ii_b5fd1b8f-en

OECD. (2025a). *Monitoring and Assessing the Impact of National Action Plans Against Racism*. OECD Publishing. https://doi.org/10.1787/13e32676-en

OECD. (2025b). OECD Science, Technology and Innovation Outlook 2025: Driving Change in a Shifting Landscape. *OECD Science, Technology and Innovation Outlook*, *2025*. https://doi.org/10.1787/5fe57b90-en

Ohanian, R. (1990). Construction and Validation of a Scale to Measure Celebrity Endorsers' Perceived Expertise, Trustworthiness, and Attractiveness. *Journal of Advertising*, *19*(3), 39–52. https://doi.org/10.1080/00913367.1990.10673191

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*(2), 171–192. https://doi.org/10.1037/a0032734

Paolini, S., Harwood, J., Rubin, M., Huck, J., Dunn, K., & Dixon, J. (2024). Reaching across social divides deliberately: Theoretical, political, and practical implications of intergroup contact volition for intergroup relations. *Social and Personality Psychology Compass*, *18*(8), e12988. https://doi.org/10.1111/spc3.12988

Park, J., Felix, K., & Lee, G. (2007). Implicit Attitudes Toward Arab-Muslims and the Moderating Effects of Social Information. *Basic and Applied Social Psychology*, *29*(1), 35–45. https://doi.org/10.1080/01973530701330942

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*(3), 277–293. https://doi.org/10.1037/0022-3514.89.3.277

*Pech, G., & Caspar, E. (2024). Can a Video Game with a Fictional Minority Group Decrease Intergroup Biases towards Non-Fictional Minorities? A Social Neuroscience Study. *International Journal of Human-Computer Interaction*, *40*(2), 482–496. APA PsycInfo. https://doi.org/10.1080/10447318.2022.2121052

*Peña, J., & Hernández Pérez, J. F. (2020). Game perspective-taking effects on willingness to help immigrants: A replication study with a Spanish sample. *New Media & Society*, *22*(6), 944–958. APA PsycInfo. https://doi.org/10.1177/1461444819874472

*Peña, J., Pérez, J., Khan, S., & Gómez, A. (2018). Game Perspective-Taking Effects on Players' Behavioral Intention, Attitudes, Subjective Norms, and Self-Efficacy to Help Immigrants: The Case of "Papers, Please." *Cyberpsychology, Behavior, and Social Networking*, *21*(11), 687–693. https://doi.org/10.1089/cyber.2018.0030

Peng, W., Lee, M., & Heeter, C. (2010). The Effects of a Serious Game on Role-Taking and Willingness to Help. *Journal of Communication*, *60*(4), 723–742. https://doi.org/10.1111/j.1460-2466.2010.01511.x

Pettigrew, T. F., & Meertens, R. W. (1995). Subtle and blatant prejudice in western Europe. *European Journal of Social Psychology*, *25*(1), 57–75. https://doi.org/10.1002/ejsp.2420250106

Playfish. (2011). *The Sims Social* [Computer software]. Electronic Arts. https://en.wikipedia.org/w/index.php?title=The_Sims_Social&oldid=1303345527

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, *67*(4), 741–763. https://doi.org/10.1037/0022-3514.67.4.741

Priego-Ojeda, M., Filella-Guiu, G., & Pérez-Escoda, N. (2025). Playing With Emotions: Evaluating a Serious Game to Promote Parents' Emotional Competencies Within the Family Context. *Journal of Computer Assisted Learning*, *41*(5), e70126. https://doi.org/10.1111/jcal.70126

Qaffas, A. A. (2020). An Operational Study of Video Games' Genres. *International Journal of Interactive Mobile Technologies (iJIM)*, *14*(15), 175–194. https://doi.org/10.3991/ijim.v14i15.16691

Ramasubramanian, S. (2010). Television viewing, racial attitudes, and policy preferences: Exploring the role of social identity and intergroup emotions in influencing support for affirmative action. *Communication Monographs*, *77*(1), 102–120. https://doi.org/10.1080/03637750903514300

Ramasubramanian, S. (2011). The Impact of Stereotypical Versus Counterstereotypical Media Exemplars on Racial Attitudes, Causal Attributions, and Support for Affirmative Action. *Communication Research*, *38*(4), 497–516. https://doi.org/10.1177/0093650210384854

Rosenthal, L., & Levy, S. R. (2010). The Colorblind, Multicultural, and Polycultural Ideological Approaches to Improving Intergroup Attitudes and Relations. *Social Issues and Policy Review*, *4*(1), 215–246. https://doi.org/10.1111/j.1751-2409.2010.01022.x

*Saleem, M., & Anderson, C. (2013). Arabs as Terrorists: Effects of Stereotypes Within Violent Contexts on Attitudes, Perceptions, and Affect. *Psychology of Violence*, *3*(1), 84–99. APA PsycInfo. https://doi.org/10.1037/a0030038

Salter, P. S., Adams, G., & Perez, M. J. (2018). Racism in the Structure of Everyday Worlds: A Cultural-Psychological Perspective. *Current Directions in Psychological Science*, *27*(3), 150–155. https://doi.org/10.1177/0963721417724239

Saunders, C., Rutkowski, A. F., Genuchten van, M., Vogel, D., & Orrego, J. M. (2011). Virtual Space and Place: Theory And Test. *MIS Quarterly*, *35*(4), 1079–1098. https://doi.org/10.2307/41409974

Schwarzenthal, M., Schachner, M. K., Juang, L. P., & van de Vijver, F. J. R. (2020). Reaping

the benefits of cultural diversity: Classroom cultural diversity climate and students' intercultural competence. *European Journal of Social Psychology*, *50*(2), 323–346. https://doi.org/10.1002/ejsp.2617

Shen, L. (2010). On a Scale of State Empathy During Message Processing. *Western Journal of Communication*, *74*(5), 504–524. https://doi.org/10.1080/10570314.2010.512278

Srivastava, S., Tamir, M., McGonigal, K. M., John, O. P., & Gross, J. J. (2009). The social costs of emotional suppression: A prospective study of the transition to college. *Journal of Personality and Social Psychology*, *96*(4), 883–897. https://doi.org/10.1037/a0014755

Stephan, W. G., & Stephan, C. W. (1985). Intergroup Anxiety. *Journal of Social Issues*, *41*(3), 157–175. https://doi.org/10.1111/j.1540-4560.1985.tb01134.x

Strahan, R., & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlow-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, *28*(2), 191–193. https://doi.org/10.1002/1097-4679(197204)28:2%253C191::AID-JCLP2270280220%253E3.0.CO;2-G

Sue, D. W., Capodilupo, C. M., Torino, G. C., Bucceri, J. M., Holder, A. M. B., Nadal, K. L., & Esquilin, M. (2007). Racial microaggressions in everyday life: Implications for clinical practice. *American Psychologist*, *62*(4), 271–286. https://doi.org/10.1037/0003-066X.62.4.271

Tamborini, R., Ron, Eastin, M., S., M., Skalski, Paul, Lachlan, K., Kenneth, Fediuk, A., T., Brady, & Robert. (2004). Violent Virtual Video Games and Hostile Thoughts. *Journal of Broadcasting & Electronic Media*, *48*, 335–357.

Tan, A., Fujioka, Y., & Tan, G. (2000). Television use, stereotypes of African Americans and opinions on affirmative action: An affective model of policy reasoning. *Communication Monographs*, *67*(4), 362–371. https://doi.org/10.1080/03637750009376517

Tan, A., Fujioka, Y., & Tan, G. (2009). Television use, stereotypes of African Americans and opinions on affirmative action: An affective model of policy reasoning. *Communication Monographs*, *67*(4), 362–371. https://doi.org/10.1080/03637750009376517

Teichman, Y., Bar-Tal, D., & Abdolrazeq, Y. (2007). Intergroup biases in conflict: Reexamination with Arab pre-adolescents and adolescents. *International Journal of Behavioral Development*, *31*(5), 423–432. https://doi.org/10.1177/0165025407081470

The Pixel Hunt. (2019). *Bury Me, My Love* [Computer software]. PID Games. https://store.steampowered.com/app/808090/Bury_Me_My_Love/

Thompson, M. M., Naccarato, M. E., Parker, K. C. H., & Moskowitz, G. B. (2001). The personal need for structure and personal fear of invalidity measures: Historical perspectives, current applications, and future directions. In *Cognitive social psychology: The Princeton Symposium on the Legacy and Future of Social Cognition* (pp. 19–39). Lawrence Erlbaum Associates Publishers.

Treyarch. (2010). *Call of Duty: Black Ops* [Computer software]. Activision. https://store.steampowered.com/app/42700/Call_of_Duty_Black_Ops/

Tropp, L. R., White, F., Rucinski, C. L., & Tredoux, C. (2022). Intergroup Contact and Prejudice Reduction: Prospects and Challenges in Changing Youth Attitudes. *Review of General Psychology*, *26*(3), 342–360. https://doi.org/10.1177/10892680211046517

US Census Bureau. (2019). *Projections of the Size and Composition of the U.S. Population: 2014 to 2060*. Census.Gov. https://www.census.gov/library/publications/2015/demo/p25-1143.html

Valve. (2000). *Counter-Strike* [Computer software]. Sierra Studios. https://store.steampowered.com/app/10/CounterStrike/

Van Dyne, L., Ang, S., Ng, K. Y., Rockstuhl, T., Tan, M. L., & Koh, C. (2012). Sub-Dimensions of the Four Factor Model of Cultural Intelligence: Expanding the Conceptualization and Measurement of Cultural Intelligence. *Social and Personality Psychology Compass*, *6*(4), 295–313. https://doi.org/10.1111/j.1751-9004.2012.00429.x

Van Looy, J., Courtois, C., De Vocht, M., & De Marez, L. (2012). Player Identification in Online Games: Validation of a Scale for Measuring Identification in MMOGs. *Media Psychology*, *15*(2), 197–221. https://doi.org/10.1080/15213269.2012.674917

*Vang, M., & Fox, J. (2014). Race in Virtual Environments: Competitive Versus Cooperative Games with Black or White Avatars. *Cyberpsychology, Behavior, and Social Networking*, *17*(4), 235–240. https://doi.org/10.1089/cyber.2013.0289

Visual Concepts. (2013). *NBA 2K14* [Computer software]. 2K. https://steamcommunity.com/app/255480?l=german

Volition. (2008). *Saints Row 2* [Computer software]. THQ. https://store.steampowered.com/app/9480/Saints_Row_2/

Wanako Games. (2006). *3D Ultra Minigolf Adventures* [Computer software]. Vivendi Games. https://en.wikipedia.org/w/index.php?title=3D_Ultra_Minigolf_Adventures&oldid=1305241979

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240. https://doi.org/10.1162/105474698565686

*Yang, G., Gibson, B., Lueke, A., Huesmann, L., & Bushman, B. (2014). Effects of Avatar Race in Violent Video Games on Racial Attitudes and Aggression. *Social Psychological and Personality Science*, *5*(6), 698–704. APA PsycInfo. https://doi.org/10.1177/1948550614528008

Yuke's Tose. (2009). *WWE SmackDown vs. Raw 2010* [Computer software]. THQ. https://en.wikipedia.org/w/index.php?title=WWE_SmackDown_vs._Raw_2010&oldid=1302687447

Zavala-Rojas, D. (2014). Thermometer Scale (Feeling Thermometer). In *Encyclopedia of Quality of Life and Well-Being Research* (pp. 7199–7200). Springer, Cham. https://doi.org/10.1007/978-3-031-17299-1_1028

**Appendix A**

*Studies Focusing on the Effects of Collaborative Gameplay: Characteristics of the Sample, the Video game, Study Design, Outcome Variables, Analytical Approach and Main Findings*

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| *Adachi et al. (2015)* | | | | |
| $N = 138$ Canadian university students, 56% female, $M_{age} = 18.75$ ($SD = 2.17$) | *Call of Duty: Black Ops* [FPS]<br><br>A fast-paced military shooter played from a first-person perspective. Players engage in armed combat using firearms, explosives, and tactical equipment, either completing narrative missions in single-player or competing in team-based multiplayer matches. | Two students participated together in different rooms; they were told that the study was about personality and video gameplay. Participant interacted live using XBOXLive.<br>In the intergroup condition, participants were told that they were cooperating online with American students from Buffalo University. The intragroup condition participants were told the truth about who they were playing with (i.e., other Canadian students from Brock University). Participants could not communicate. Every participant completed a demographic and pre-test intergroup attitude questionnaire, played the video game for 12 minutes, filled in a post-test attitude measures and a discrimination task. | *Prejudice:* Attitudes toward Americans (national outgroup) and University of Buffalo students (university outgroup) using feeling thermometers and attitudes toward other outgroups (e.g., the elderly, alcoholics, immigrants) were assessed.<br>*Categorization as one team:* Feeling of being "one team" at post-test (see Gaertner & Dovidio, 2013).<br>*Discrimination:* Selecting a video for their partner to watch, eliciting either negative, neutral, or positive feelings at post-test (adapted from Hodson et al., 2003). | Results from repeated measures ANOVA showed that attitudes toward both Americans and University of Buffalo students became significantly more favourable after cooperating with an outgroup member, with large effect sizes (Cohen's $d$ of 1.39 and 1.15). By contrast, participants who cooperated with an ingroup member exhibited no changes in attitudes. Further, postgame attitudes toward outgroup members (Americans, University of Buffalo students) were significantly more favourably among participants of the intergroup (vs. intragroup) condition, with medium effect sizes (Cohen's $d$ of 0.52 and 0.54).<br>There were no secondary transfer effects to changes in attitudes toward other outgroups (e.g., the elderly, immigrants). Both in the intra- and intergroup condition, participants categorized themselves and their cooperation partner as one team. Finally, intragroup participants did not discriminate against their outgroup partner, and there were no significant differences in their tendencies to discriminate a (future) ingroup member versus their current outgroup member. |
| *Adachi et al. (2016)* | | | | |
| Experiment 1: $N = 77$ Canadian university students, 67% | Experiment 1: *Call of Duty: Black Ops* (see Adachi et al., 2015) | Experiment 1: Same procedure as in Adachi et al. (2015) but here, participants either played the game cooperatively with an outgroup member (American University of Buffalo student) or with an | Experiments 1 and 2 used the same measures as in Adachi et al. (2015) with one addition: Empathy toward an outgroup | Experiment 1: Results from repeated measures ANOVA showed that participants in the intergroup cooperation context reported more favourable post-test outgroup attitudes, representing a large effect (Cohen's $d$ =1.28). Playing the video game alone did not result in any outgroup attitude changes. As in the 2015-study, no evidence of a |

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| female $M_{age}$ = 18.7<br><br>Experiment 2:<br>$N$ = 239 Canadian university students, 66% female, $M_{age}$ = 19.08 ($SD$ = 1.17) | Experiment 2: *Call of Duty: Black Ops*<br>(see above)<br><br>*NBA 2K14*<br>[Sports Game]<br>A basketball simulation game Players control real-life teams and athletes in matches, focusing on teamwork, strategy, and individual performance. Gameplay involves passing, shooting, defence, and coordinated team play in both competitive and cooperative modes. | ingroup member (Canadian University of Brock student, as in Adachi et al., 2015) or alone (solo-play mode).<br>Participants were told that they were in two unrelated studies on relations between personality and video gameplay and on impressions formed of other university students' experience (to disguise purpose of the empathy measure).<br>Participants provided outgroup attitude ratings before playing the video game for 12 minutes and then completed the other measures.<br><br>Experiment 2: The experiment was conducted exactly like the first one, with one exception: Participants were randomly assigned to play either a violent (*Call of Duty: Black Ops*) or nonviolent (*NBA 2K14*) video game, either cooperatively or alone. | participant (Greitemeyer, 2013). | secondary transfer to unrelated outgroups was found. Further ANOVA results revealed that students in the intergroup cooperation versus solo-play condition were more likely to select positive videos for the outgroup participant to view (discrimination task). Mediation analyses showed a significant effect, with participants in the cooperation (vs. solo) condition reporting greater recategorization, which, in turn, predicted more favourable post-test outgroup attitude.<br><br>Experiment 2: Results from repeated measures ANOVA showed that effects did not differ violent and nonviolent video games; therefore, in subsequent analyses, the two groups were collapsed into general intergroup cooperation versus solo-play. Attitudes toward the outgroup became more favourable in both conditions; however, the effect of intergroup cooperation from pre- to post-game was twice the size (Cohen's $d$ = 1.63) of playing alone (Cohen's $d$ = .81). There was no significant effect for the discrimination task, and the effect of intergroup cooperation on discrimination did not differ as a function of playing the violent versus the nonviolent video game. Finally, mediation analyses revealed a significant indirect effect of context on post-test outgroup attitudes via recategorization. Participants in the cooperation (vs. solo) context reported greater recategorization, which, in turn, predicted more favourable post-test outgroup attitudes. No indirect effect of empathy emerged. In addition, there was a significant indirect effect of context on discrimination via recategorization, but not via empathy. |
| *Benatov et al. (2021)* | | | | |
| $N$ = 89 6th-grade students. | *Minecraft*<br>[Sandbox Games] | In both groups, *Minecraft* was played for 90 minutes six times. In each session, there was a challenge that had to be solved as teams. | *Emotional prejudice*: Rating six different emotions toward members of the | Six of the seven constructs showed an interaction between the factors intervention and time, with large effects for emotional prejudice against outgroup ($\eta_p^2$ = .258), willingness for social contact with outgroup ($\eta_p^2$ = .200), |

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| Experimental group $N = 46$ (52.5 % female, 50 % Jewish-Israeli, 50 % Palestinian-Israeli);<br>Control group $N = 43$ (50.4 % female, 51.1 % Jewish-Israeli, 48.9 % Palestinian-Israeli) | A sandbox game without a set storyline where players explore, build, and survive in procedurally generated block-based worlds. Core mechanics include resource gathering, crafting, and construction, with both cooperative and competitive multiplayer modes enabling collaborative projects or challenges. | Students were encouraged to take part in "bonus" challenges after school. The teams competed against each other over the course of the program.<br>In the *treatment group*, teams were mixed between Jewish-Israeli and Palestinian-Israeli students. The *control group* played only with peers in their class. A text chat with real-time translation between Arabic and Hebrew could be used for communication, but participants were free to use other communication platforms. In addition to the six online meetings, two real meetings took place after the third and sixth session, where the Jewish-Israeli and Palestinian-Israeli students in the treatment group visited each other.<br>Prejudice toward the other group was assessed in a pre-test one week before intervention, a post-test immediately after, and a follow-up test at six months. | other group (Teichman et al., 2007).<br>*Positive and negative characteristics*: Naming all characteristics of the other group that they could think of. The number of positive and negative characteristics was counted.<br>*Negative stereotyping*: Rating ten traits for their own and the other group (Kaminsky & Bar-Tal, 1996).<br>*Negative behavioural expectations*: Probability of a new neighbour of the other group showing eight positive/negative behaviours (Berger et al., 2018).<br>*Willingness for social contact*: Willing to perform eight activities with members of the other group (Teichman et al., 2007).<br>*Actual contact*: Frequency with which they interact with the other group. | negative characteristics of outgroup ($\eta_p^2 = .156$,), negative outgroup stereotyping ($\eta_p^2 = .156$) and medium effects for positive characteristics of outgroup ($\eta_p^2 = .137$) as well as negative behavioural expectations of outgroup ($\eta_p^2 = .128$). Only actual contact with outgroup was not significant.<br>Within-group t-tests showed medium-sized significant differences for six of seven variables between the first and second measurement time points as well as between the first and third measurement time points in the treatment group: willingness for social contact (1 to 2: $d = -0.64$; 1 to 3: $d = -0.68$), negative outgroup stereotyping (1 to 2: $d = 0.59$; 1 to 3: $d = 0.61$), emotional prejudice against outgroup (1 to 2: $d = 0.52$; 1 to 3: $d. = 0.56$), positive characteristics of outgroup (1 to 2: $d = 0.52$; 1 to 3: $d = 0.56$), and negative behavioural expectations (1 to 2: $d = 0.50$; 1 to 3: $d = 0.52$). Negative characteristics of outgroup was not significant. The observed effects remained stable at follow-up.<br>In the control group, prejudice increased significantly in two of seven variables between the first and second measurement time points, with a medium effect for negative outgroup stereotyping (1 to 2: $d = 0.59$; 1 to 3: $d = 0.61$) and a small effect for emotional prejudice against outgroup (1 to 2: $d = -0.23$; 1 to 3: $d = -0.23$). The authors hypothesize these findings might be due to a murder at Haifa University that year. There were no significant changes between post-test and follow-up. |

**Appendix B**

*Studies Focusing on the Effects of Playing or Interacting with Different Game Characters: Characteristics of the Sample, the Video game, Study Design, Outcome Variables, Analytical Approach and Main Findings*

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| *Alhabash & Wise (2012)* | | | | |
| *N* = 68 university students from the US, 74% female, $M_{age}$ = 20; however, the analyses relied on 58 responses as 10 outliers were excluded. | *PeaceMaker* [Simulation and Social Games]<br><br>A political simulation game where players take the role of either the Israeli Prime Minister or the Palestinian President. Players make policy decisions, respond to real-world inspired events, and balance domestic and international pressures, with the goal of achieving a peaceful resolution to the Israeli–Palestinian conflict. | Random assignment to two experimental groups (playing the Israeli prime minister or the Palestinian president). The pre-test included explicit and implicit attitude measures. After the pre-test, participants played *PeaceMaker* for 20 minutes. The post-test included the same measures as the pre-test. | *National attitudes toward Palestinians and Israelis* (explicit attitudes) from ADL (2004) and Mayer (2004): Rating each group on multiple dimensions, such as favourability, sympathy, being democratic, being responsible for the violence etc.<br>*Implicit attitudes*: Affective misattribution task (Payne et al., 2005): Participants were exposed to portrait pictures of Israelis and Palestinians, followed by unrelated Chinese pictographs, and had to rate the pictographs as quickly as possible as either pleasant or unpleasant. | Results from repeated measures ANOVAs showed a significant 2 nationality assignment (Israelis versus Palestinians) x 2 national group (attitudes toward Israelis versus Palestinians) x 2 time (pre-test versus post-test) interaction of $\eta_p^2$ = .08: Participants who played the Israeli Prime Minister, did not show a significant change in their favourability toward Israelis or Palestinians. On the other hand, those playing the Palestinian president observed a significant decrease in their favourability of Israelis, and a significant increase in their favourability of Palestinians as a function of playing *PeaceMaker*. No significant effects were observed for implicit attitudes. |
| *Alhabash & Wise (2014)* | | | | |
| *N* = 172 university students from | *PeaceMaker* (see Alhabash & Wise, 2012). | Largely the same procedure as in Alhabash & Wise (2012). | *Explicit stereotypes regarding Palestinians and Israeli* (Bar-Tal & Labin (2001). | Results from paired-samples *t*-Tests revealed that participants expressed significantly more positive stereotypes about Israelis than Palestinians (Cohen's *d* = |

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| the US, 60.4% female, 75.6% Caucasian/White, $M_{age}$ = 20.8, *SD* = 2.89, 14 cases discarded (i.e., analyzed *N* = 158). | | | *Implicit attitudes* (see Alhabash & Wise (2012), above). | .43) and had significantly more positive implicit attitudes toward Israelis than Palestinians (Cohen's *d* = .28). Results from repeated measures analyses of variance showed a significant interaction between role assignment, national group, and time ($\eta_p^2$ = .05): Participants who played the role of the Palestinian President (Israeli Prime Minister) showed a significant negative change in their explicit stereotypes about Israelis (Palestinians) from pre- to post-test, whereas their explicit stereotypes about Palestinians (Israelis) did not change significantly. There were no significant effects on changes in implicit attitudes. |
| Ash (2016) | | | | |
| *N* = 84 university students from the US ($M_{age}$ = 20.93, 60.7% female, 85.7% White, 7.1% Asian, 3.6% Hispanic or Latino, and 3.6% African American). | *Fight Night Round 4* [Sports Game]<br><br>A boxing simulation game where players control custom-created or real-life fighters such as Muhammad Ali. Fast-paced matches emphasize punching mechanics, stamina management, and strategy. | Participants were randomly assigned to a 2 avatar race (White or Black) x 2 opponent race (White or Black) experimental design.<br>Prior to playing, the participants filled in a demographic questionnaire. After completing a tutorial of the video game, they played two 3-minutes rounds and completed questionnaires. | *Aggressive cognition:* Word completion task (Anderson et al., 2003; Eastin et al., 2009).<br>*Aggressive affect* (Anderson et al., 1995).<br>*Aggressive behaviour*: Punching the opponent in the game<br>*Stereotypes:* (describing the avatar, Kawakami & Dovidio (2001)<br>*Avatar embodiment*: Identification with the avatar | Results from regression analyses showed that avatar race (Black or White) did not predict Black or White stereotype use (but, regardless of race, the boxing avatars were describes as having more Black stereotypes than White stereotypes). Further avatar race did not predict in-game aggressive behaviour, post-game cognition, or post-game affect. It was further tested whether Black stereotype use functions as a moderator of the effects of avatar race on the outcomes in that the positive relation between using a Black avatar and aggression is stronger for players describing the avatar using stereotypes. None of the interactions was significant. Use of Black stereotypes to describe one's avatar negatively predicted aggressive thoughts, and stereotype use was associated with increased levels of aggressive affect. Avatar race was a significant predictor of aggressive in-game behaviour when participants experienced levels of embodiment greater than 1 *SD* above the mean of embodiment. No significant interactions (avatar race x embodiment) occurred for cognition and affect. Finally, playing as a Black avatar against a White opponent was not associated with higher levels of outcome variables. |

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| Behm-Morawitz, Hoffswell, et al. (2016) | | | | |
| $N$ = 107 undergraduate university students from the US ($M_{age}$ = 20, $SD$ = 1.38; 54.2 % female, 100 %, White)<br><br>Experimental group ($n$ = 53, 60.4 % female);<br>Control group ($n$ = 54, 48.1 % female) | *Saints Row 2* [Third-person Action and Adventure Games]<br><br>An open-world action-adventure game where players take the role of a customizable gang member navigating a fictional city. Gameplay combines story-driven missions with free exploration, including combat, driving, and various side activities, often featuring violent and criminal themes. | Participants were randomly assigned to either the gaming condition or the control condition.<br>Participants in the gaming condition played Saints Row 2 for 30 Minutes. During the gameplay-sequence, players attempted to escape a prison. Male participants were assigned an African American male character, and female participants were assigned an African American female character. After playing, the participants answered an online questionnaire.<br>The control group answered the questionnaire directly without playing the game.. | *Stereotyping*:<br>1) Beliefs about African Americans as a difference score between the means of beliefs about African Americans and Whites.<br>2) Race-related policy beliefs (Ramasubramanian, 2010) measured self-reported support for policies aimed at reducing inequality. For both measures, higher scores indicate higher negative stereotyping.<br>*Emotional response*: Positive and Negative Affect Scale (PANAS; Watson et al., 1988).<br>*Spatial presence*: Presence Questionnaire (Witmer & Singer, 1998) [assessed only in the intervention group] | The analyses showed that negative affect predicted negative beliefs about African Americans ($b$ = -0.20, $p$ = .05, 95%CI [-0.4306, -0.0003]) and lower policy support for minorities ($b$ = -0.62, $p$ = .01). The authors conclude that playing reinforces stereotyping through the virtual threat effect, which is an adoption of the incremental threat effect (Crisp et al., 2009). No significant results were found for the second assumption, that spatial presence in the game reinforced stereotyping. There were no significant moderating effects for character/player gender on negative affect or stereotyping. For the control group, no significant associations between the variables were found. An ANOVA showed that playing Saints Row 2 predicted significant fewer positive beliefs about African Americans than the control group ($R^2$ = .12), but there was no significant prediction of race-related policy beliefs. |
| Behm-Morawitz, Pennell, et al. (2016) | | | | |
| $N$ = 87 undergraduate students from the US ($M_{age}$ = 21, $SD$ = 1.20, | *The Sims Social* [Simulation and Social Games] | Participants were asked to create an avatar and play the sims social for 30 minutes. Randomly, participants were assigned to create an avatar with a Black or | *Self-presence*: Extent to which the avatar is perceived as an extension of oneself (Behm-Morawitz, 2013). | The authors assessed whether the effects of avatar skin colour on race-related policy beliefs is mediated by beliefs about African American women/men.<br>Playing a Black avatar resulted in more favourable beliefs about male African Americans ($b$ = 1.87, $SE$ = 0.82), but |

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| 37 % female, 100 % White)<br><br>White avatar condition ($n$ = 42);<br>Black Avatar condition ($n$ = 45) | A social version of The Sims series, integrated with Facebook, where players create and manage avatars, build homes, and interact with others in a virtual environment. Gameplay emphasizes everyday activities and social relationships, allowing players to explore identity and interpersonal dynamics. | White skin tone. The gender of the avatar was matched to the participants' gender. After playing, the participants answered an online questionnaire. | *Avatar liking*: Feelings of warmth and liking toward the in-game avatar.<br>*Race-related beliefs*: Participants performed an Implicit Association test (based on Nosek, n.d.) by assigning 8 positive or negative characteristics to pictures of five faces (African American Female, African American Male, White Female, White Male, Ambiguous Race Male [distractor]).<br>*Race-related policy beliefs*: Self-reported support for policies aimed to reduce inequality (adapted from Tan et al. (2009); Ramasubramanian (2011). | not about female African Americans. Regardless of whether beliefs about male or female African Americans were used as mediator variable, there was a significant direct effect of avatar skin tone on race-related minority policy beliefs (male mediator variable: $b$ = 0.47, $SE$ = 0.19; female mediator variable: $b$ = 0.55, $SE$ = 0.20. The effects of avatar skin tone on race-related policy beliefs were not mediated by beliefs about male or female African Americans. The gender of the participants did not moderate the effects of avatar skin tone on beliefs about male or female African Americans.<br>Further, it was considered how self-presence in the avatar and avatar liking moderated the association of avatar skin with race-related policy beliefs and beliefs about male and female African Americans. For self-presence, there was neither significant direct effects, nor significant moderation effects. For avatar liking, there was a significant interaction effect for race-related policy beliefs ($b$ = 0.63, $SE$ = 0.20), with high avatar liking being associated with higher pro-minority policy beliefs. There was no interaction effect for beliefs about male or female African Americans. |
| *Breves (2020)* | | | | |
| $N$ = 86 undergraduate and graduate Students from Germany ($M_{age}$ = 20.98, $SD$ = 1.77, 50 % female, 100% White) | *The Elder Scrolls V: Skyrim* [Third-person Action and Adventure Games]<br><br>An open-world role-playing game where players explore a vast fantasy world, complete quests, | All players were asked to complete a quest in *The Elder Scrolls V: Skyrim* in three conditions. Participants in the 2D condition and in the VR condition were given the task by a Black non-playable character (NPC). In the control condition, a research assistant gave the task in the same wording instead. The game was modified to exclude violent confrontations with creatures. | *Spatial presence*: The Spatial Presence Experience Scale by Hartmann et al. (2016) was used to assess perceived self-location presence and possible actions in the game.<br>*Levels of explicit prejudice*: Participants explicit prejudice toward Black persons was measured in three dimensions (threat | A one-way ANOVA showed a significant medium effect for explicit prejudice between groups ($\eta^2$ = .07). However, simple contrasts showed that the explicit prejudice was only lower in the VR condition compared to the control condition (Contrast = 0.37).<br>There was no significant difference among the three groups for implicit prejudice.<br>An ANCOVA showed a significant medium-sized effect for spatial presence ($\eta^2$ = .08) in the VR condition ($M$ = 4.79, $SD$ = 1.08) compared to the 2D condition ($M$ = 4.07, $SD$ = 1.09) and control condition ($M$ = 4.25, $SD$ = 1.19). |

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| Control condition ($n$ = 27);<br>2D condition ($n$ = 29);<br>VR condition ($n$ = 30) | and interact with a wide range of characters. The game can be played in first- or third-person perspective and emphasizes exploration, character development, and immersive storytelling. | | and rejection, intimacy and traditional values; Pettigrew & Meertens, 1995).<br>*Levels of implicit prejudice*: An implicit association test was used (based on Greenwald et al., 2003). The test measures whether participants associate negative concepts with Black more quickly than with Whites and vice versa for positive terms. | |
| *V. Chen et al., (2025)* | | | | |
| $N$ = 300 ($M_{age}$ = 22.77, $SD$ = 3.15; 65.3% female), from Singapore<br><br>Moral choices + all citizens ($n$ = 50); Moral choices + citizens and foreigners ($n$ = 50) Moral choices + all foreigners ($n$ = 50); non-moral choices + all citizens ($n$ = 51); non-moral | *Hire Up* [Self-developed Games]<br><br>A self-developed management simulation where players act as company managers making hiring and workplace decisions. In the moral condition, players faced six scenarios where they could either protect workers (losing credits) or exploit them | The study was conducted online. Three to five days before playing the game, participants completed a pre-game questionnaire on their attitudes toward immigrants and trait moral foundations. Participants were randomly assigned to one of six conditions in a 2 presence of moral conflicts (moral choices or non-moral choices) x 3 the nationality of NPCs (all foreigners or all citizens or citizens and foreigners) design. After playing Hire Up, participants indicated their attitudes toward immigrants and their guilt experienced while playing. | *Attitudes toward immigrants*: Valence of attitudes toward immigrants was measured with the feeling thermometer on a scale from 0 to 100 (Zavala-Rojas, 2014).<br>*Moral foundations*: Five trait moral foundations (care/harm, fairness/reciprocity, ingroup/loyalty, authority/subversion, purity/sanctity and liberty/oppression) were measured using the Moral Foundations Questionnaire (Graham et al., 2008). | Results of an *ANCOVA* showed that the presence of moral decisions did not influence post-game attitudes toward immigrants. The nationality of NPCs did have a small significant effect on post-game attitudes toward immigrants ($\eta_p^2$ = .02.). Post-hoc comparisons revealed that this effect was only significant for participants who interacted with Citizens and Foreigners reporting lower attitudes toward immigrants than participants who interacted with only Citizens. The presence of moral choices and NPC nationality did not interact significantly.<br>A *t*-test revealed that participants playing the game with moral decisions experienced significantly more guilt than participants playing the game with non-moral decisions.<br>A path analysis was performed on the subsample of participants who had interacted with foreigner NPCs in the game version containing moral choices ($n$ = 100). The analysis revealed that participants reporting higher trait moral foundation in fairness/reciprocity were significantly less likely to violate fairness/reciprocity (β = .28, $SE$ =.12). |

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| choices + citizens and foreigners ($n$ = 50); non-moral choices + all foreigners ($n$ = 49) | (gaining credits). In the non-moral condition, players encountered parallel scenarios without moral implications. NPC well-being and players' in-game credits changed depending on their choices, with feedback provided both during and after gameplay. | | *Experienced Guilt*: Feelings of guilt experienced while playing were measured using a four-item scale adapted from Hartmann et al. (2010).<br>*Moral decisions*: In-game decisions were coded dichotomous with 0 = moral foundation upheld and 1 = moral foundation violated. | There were no similar effects for other trait moral foundations. In addition, participants who violated the care/harm (β = .28, *SE* = .09), ingroup/loyalty (β = .22, *SE* = .10), and liberty/oppression (β = .20, *SE* = .09) moral trait foundations reported higher levels of guilt. There were no similar effects for other trait moral foundations. Experienced guilt was not significantly associated with post-game attitudes toward immigrants and did not mediate the effect of moral decisions on post-game attitudes toward immigrants.<br>The authors additionally found that participants who violated the fairness/reciprocity (β = .19, *SE* = .08) und liberty/oppression (β = .16, *SE* =.07) moral foundations reported poorer attitudes toward immigrants after playing. There were no similar effects for other moral foundations. |
| *Correll et al. (2002)* | | | | |
| Experiment 1: $N$ = 40 (60% female, 39 Whites, 1 Latino) undergraduate students in the US<br><br>Experiment 2: $N$ = 44 (75% female, 42 Whites, 1 Latino, 1 Asian) undergraduate students in the US; 5 participants | *Self-developed simplistic video game* [Self-developed Games]<br><br>A custom-developed shooter task where players rapidly decided whether to "shoot" or "not shoot" male targets (Black or White) appearing in urban backgrounds while holding either a gun or a | Experiment 1: All participants played the same version of the game. Reaction times were compared in a 2 target ethnicity (Black or White) x 2 Object Type (gun or no gun) design. 7% of trials were excluded due to reaction times over 850ms. Participants with high scores received an additional monetary reward.<br><br>Experiment 2: Similar procedure as in Experiment 1. However, participants received monetary compensation based on their performance and the time window | *Reaction time*: Participants reaction time for making a correct shoot/don't shoot-decision.<br>*Error Rate*: The number of trials participants erroneously decided to shoot an unarmed target or don't shoot an armed target.<br><br>Additions in Experiment 3:<br>*Prejudice against African Americans*: Measured across different facets. The questionnaire used the Modern Racism Scale (J. B. McConahay et al., | Experiment 1: An ANOVA revealed a significant main effect for Object Type on reaction time, with participants deciding to shoot an armed target faster than deciding to not shoot an unarmed target. The interaction between object type and target ethnicity was significant, with participants shooting an armed target more quickly if he was African American and not shooting an unarmed target more quickly if he was White.<br>Regarding error rates, an ANOVA revealed a significant main effect for object type, with errors being more frequent in the no-gun condition. There was also a significant interaction between object type and target ethnicity, with falsely shooting an unarmed target being more frequent for African American targets.<br><br>Experiment 2: Data on six targets were excluded from further analysis due to particularly high error rates. There were no significant effects of object type or target ethnicity |

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| excluded for exceeding 10 timeouts<br><br>Experiment 3: *N* = 48 (65% female, 2 Latino, 1 Asian, 1 African American, 44 White), 3 participants removed from sample<br><br>Experiment 4: *N* = 52 adults from Denver (*N* = 46, 30,4% female)<br><br>African American Group (*n* = 25, 24% female)<br><br>White group (*n* = 21, 38,1% female) | non-weapon object (e.g., wallet, phone, camera). Each trial presented randomized backgrounds, and target poses to increase unpredictability. The game simulated quick decision-making under threat, designed to measure shooter bias and implicit associations between race and danger. | for responses was shortened to 630ms.<br><br>Experiment 3: Game procedure similar to Experiment 1, with timeout threshold at 850ms. After playing the game, participants answered a questionnaire. The game and questionnaire were framed to belong to different studies.<br><br>Experiment 4: Game procedure similar to Experiment 1 and 3, with timeout threshold at 850ms. Targets identified as error-prone in Experiment 2 were altered for clarification. Participant Ethnicity (African American or White) was introduced as a between-subjects factor. Participants used a laptop to participate in a non-lab setting. | 1981), the Discrimination Scale and Diversity Scale (both from Wittenbrink et al. ( 1997) and the Motivation to Control Prejudiced Responding Scale (Dunton & Fazio, 1997). Items from those scales were mixed with filler items from the Right-Wing Authoritarianism Scale (Altemeyer, 1988) and the Personal Need for Structure Scale (Thompson et al., 2001).<br>*Endorsing of negative stereotypes*: Participants were asked to rate how dangerous, violent and aggressive they perceived African Americans and Whites, based on personal beliefs. A personal stereotype index was calculated by averaging the difference score for the three items.<br>An *index for negative cultural stereotype* was created by asking participants to rate the same statements not based on their personal belief but on how they think most White Americans would estimate them. Difference scores | on reaction time. An ANOVA revealed that object type had a significant effect on error rate, with false decisions being more frequent when the target was unarmed. there was also a significant interaction of object type and target ethnicity, with wrong decisions for unarmed targets being more frequent for African American targets and wrong decisions for armed targets being more frequent for White targets.<br>Additional analyses were conducted on participants' sensitivity and response criterion for African American and White targets based on the signal detection model (Green & Swets, 1966/1974; MacMillan & Creelman, 1991). There were no significant differences in sensitivity between African American and White targets. However, there was a significant difference in response criterion, with participants tending to shoot African American targets more often when they were less threatening ($c$ = -0.24) compared to White targets ($c$ = 0.03).<br><br>Experiment 3: 6 participants were removed from further analyses because they suspected connections between the game-task and the questionnaire. Targets identified as error-prone in Experiment 2 were excluded from the analyses.<br>Similar to Experiment 1, reaction times for armed targets were significantly shorter than for unarmed targets. A significant interaction of object type and target ethnicity revealed that participants shot armed targets more quickly if they were African American and decided to not shoot unarmed targets more quickly if they were White.<br>For error rates, there was a significant interaction of object type and target ethnicity, with participants shooting unarmed targets more frequently if they were African American.<br>Additional analyses assessed whether shooter bias (faster correct responses to unarmed Whites and armed African |

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| | | | were averaged across statements.<br>*Contact with African Americans*: Participants should rate the extent of their contact with African Americans now and in the past. | Americans) was linked to the questionnaire measures. There were no significant correlations between measures of explicit prejudice or personal endorsement of negative stereotypes and shooter bias. Perception of a negative cultural stereotype ($r$ = .34) and contact with African Americans ($r$ = .38) were positively correlated with the magnitude of shooter bias.<br><br>Experiment 4: Hypotheses were tested with 2x2x2 mixed-model ANOVAs. As in the prior experiments, reaction times for armed targets were significantly faster than for unarmed targets and there was a significant interaction between object type and target ethnicity (shooter bias). There was no significant difference in shooter bias between White and African American participants.<br>There was no significant interaction for object type and target ethnicity or main effect for participant ethnicity regarding error rates. |
| *Correll et al. (2007)* | | | | |
| Experiment 1: $N$ = 70 non-Black undergraduate students<br>Experiment 2: $N$ = 92 non-Black undergraduate students (58% female, $M_{age}$ = 20.4)<br><br>Experiment 3: $N$ = 75 non-Black undergraduate students (49% | *Self-developed simplistic video game* (see Correll et al., 2002). | Experiment 1: All participants played the same version of the game. Reaction times were compared in a 2 target ethnicity (Black or White) x 2 object type (gun or no gun) design. Timeout threshold was set to 730ms. Before playing, participants were randomly assigned to read a news article describing armed robberies about a pair of Black suspects in one condition or about a pair of White suspects in the other. They then answered distractor tasks | *Error Rate:* The amount of times participants erroneously decided to *shoot* an unarmed target or to *don't shoot* an armed target.<br><br>Experiment 2:<br>*Reaction time:* Participants reaction time for making a correct shoot/don't shoot-decision.<br>*Frequency estimate:* After each round, participants | Experiment 1: Participants had a general error rate of 7.79% across all targets. Signal detection theory analysis was used to determine how well participants were able to differentiate unarmed from armed targets (sensitivity) and how strongly they favoured shooting over not shooting (threshold). An ANOVA was used to compare values for sensitivity and threshold for both target types and article conditions. The analysis revealed that participants significantly favoured shooting over not shooting if the target was Black ($M_c$ = -.21) than when the target was White ($M_c$ = -.04). Target ethnicity interacted significantly with article condition, resulting in participants reading an article about Black criminal suspects showing stronger bias for shooting Black targets. However, participants who read an article about White criminals showed a comparable |

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| female, $M_{age}$ = 19.53) | | (e.g., memory about the news article) before playing.<br><br>Experiment 2: Each participant played two rounds. In round one, the simulation was manipulated, so that either Black targets were now more likely to be armed compared to White targets (stereotype-congruent condition) or vice-versa (stereotype-incongruent condition). Participants were randomly assigned to either condition. In round two, participants played the original version of the simulation. Timeout threshold was set to 850ms.<br><br>Experiment 3: Similar to Experiment 2, participants first were randomly assigned to play either the game-version that did not reflect any bias (control condition), the stereotype-congruent or the stereotype-incongruent version of the game. Afterwards, they completed an Extrinsic Affective Simon Task (EAST). | estimated the frequency of (un)armed Black and White targets in the game. Participants estimates were transformed to percentages.<br><br>Experiment 3:<br>*Reaction time for EAST:* Log transformed reaction time to make a correct response to a stimulus in the EAST. Participants first classified a set of male names as either African American or Caucasian by pressing two corresponding buttons. In a second trial, participants should use the two same buttons to indicate if words on a screen were either blue or green. The target-words were positive (e.g., rainbow) or negative (e.g., filth). Separate averages were computed for congruent responses and incongruent responses.<br>*Stereotype accessibility* was computed by subtracting the congruent from the incongruent average. | threshold for both Black and White targets. The article condition impacted participants threshold only for White targets after reading about White criminals. Here, participants lowered their criterion for White targets significantly.<br><br>Experiment 2: After round one, ANOVAs found participants to perceive significantly stronger associations between Black targets and weapons in the stereotype-congruent compared to the stereotype-incongruent condition. There were no main effects of condition on frequency estimate in round two. A measure of bias reflecting the interaction of Target Ethnicity and Object Type was computed by averaging the log-transformed reaction times from correct game trials. Contrast analyses revealed that Bias was significantly higher in the stereotype-congruent than in the stereotype-incongruent condition. This difference in bias was stable between both rounds.<br><br>Experiment 3: Contrast analyses revealed that stereotype accessibility significantly varied between conditions. A linear effect indicated that stereotypes were more accessible in the stereotype-congruent condition compared to the stereotype-incongruent condition. Participants in the stereotype-congruent condition responded more quickly to danger-words if they required the Black-name button. There was no effect on reaction time in the stereotype-incongruent condition. Concerning target-words not related to danger, participants in general were quicker to classify words with negative valence (e.g., filth) if they used the Black-name button. This effect did not differ significantly between conditions. |

*Eastin et al. (2009)*

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| $N$ = 178 undergraduate students from the US (53.4 % Black, 50 % female, $M_{age}$ = 22, $SD$ = 3.90). The manipulation check excluded 16 cases. | *Unreal Tournament: Game of the Year Edition* [FPS]<br><br>A classic arena-style FPS featuring fast-paced combat in enclosed maps. Players compete against human or AI-controlled opponents using a wide range of futuristic weapons and power-ups, with gameplay emphasizing speed, reflexes, and competitive action. | The participants were randomly assigned to one of four conditions: 2 avatar (White or Black) x2 opponent (White or Black). Participants' ethnicity (White or Black) was used as a third factor.<br>The participants read a short biography with pictures of their avatar and their opponent. Then they played the game for 15 minutes. | *Aggressive thoughts:* Word-completion task (Anderson, 1999; Anderson et al., 2004; Eastin, 2006). 50 out of 98 word fragments could be completed with aggressive words. The score is the ratio of completed aggressive words to neutral words.<br>*Identification:* Three items on how similar participants perceived themselves to be to their character (McKirnan et al., 1983). One item on identification (Appiah, 2001, 2006, 2007).<br>*Violent gameplay experience:* The participants were asked to rate their five favourite games concerning how violent they are and how often they have played these games. Similar to Tamborini et al. (2004), the ratings for frequency of use and violence were multiplied. The final score was computed by summing the scores for all five games. | Gender and violent gameplay experience were used as covariates in all subsequent ANCOVAs. There was a large significant interaction between participant race and avatar race on identification with the avatar ($\eta^2$ = .18), with Black participants reporting significantly higher identification when playing as a Black avatar ($M$ = 3.34, $SE$ = .17) than as a White avatar ($M$ = 1.93, $SE$ = .18). The identification of White participants was significantly greater when playing as a White avatar ($M$ = 2.84, $SE$ = .16) than as a Black avatar ($M$ = 2.20, $SD$ = .16).<br>There was no significant main effect for opponent race on identification, but a medium interaction effect between character race and opponent race on identification with the character ($\eta^2$ = .08). When playing as a Black avatar against a White opponent, character identification was significantly higher ($M$ = 3.14, $SE$ = .19) compared to playing against a Black opponent ($M$ = 2.48, $SE$ = .16). When playing as a White avatar against a White opponent, character identification was significantly lower ($M$ = 2.01, $SE$ = .15) compared to playing against a Black opponent ($M$ = 2.67, $SE$ = .16).<br>A small significant interaction emerged between participant race and character race for aggressive thoughts ($\eta^2$ = .03). White participants displayed higher levels of aggressive thoughts after gameplay when playing as a Black avatar ($M$ = .37, $SE$ = .02) compared to playing as a White avatar ($M$ = .29, $SE$ = .02). However, there was no significant difference in aggressive thoughts for Black participants between playing as a Black ($M$ = .34, $SE$ = .02) or a White ($M$ = .33, $SE$ = .02) avatar. There was no significant main effect found for opponent race on aggressive thoughts and no significant interaction between participant and opponent ethnicity. |

*Gjicali et al. (2020)*

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| *N* = 433 Participants recruited via MechanicalTurk (58 % female, 73.3% White, 7.9 % Black/African American, 7.9 % Asian/Asian American, 4.8 % Hispanic/Latino and 6.1 % multi-ethnic/multi-racial).<br><br>Beliefs All group (*n* = 150);<br>Beliefs End group (*n* = 133);<br>Beliefs None group (*n* = 150) | *Vox Populi* [Self-developed Games]<br><br>A self-developed educational game designed to assess cross-cultural competence. Players interact with NPCs from a fictional culture (Ustrad) based on hierarchy and collectivism. The game world responds dynamically to players' culturally appropriate or inappropriate actions. The goal is to learn local customs through three quests and ultimately succeed as Prime Minister of Ustrad. | Participants in all conditions played the same game, but they received different instructions about generating beliefs about the Ustradian culture. Participants completed three quests during the game, and their interactions with NPCs were recorded.<br><br>Participants were randomly assigned to one of three experimental conditions: beliefs all, beliefs at end, or no beliefs. Participants in the "Beliefs All" condition were prompted to enter their beliefs about the Ustradian culture in a dropdown-menu after each quest. They were provided with a list of statements about Ustradian culture and asked to rate how much they agreed or disagreed with each statement. Participants in the "Beliefs at End" condition received a similar prompt after completing the last quest. Participants in the "Beliefs None" condition did not receive any instruction about indicating beliefs.<br><br>After completing the game, participants took a quiz to assess their learning outcomes. The quiz included questions about Ustradian culture and customs that were covered in the game. | *Social exploration*: Includes time spent per quest, number of interactions with NPCs, different parts in communication with NPCs, and the amount of small talk in those conversations.<br><br>*Feedback types*: Depending on how participants interacted with the NPCs, the NPCs provided four types of feedback. The number of occurrences of each feedback type was counted, with correct feedback indicating culturally appropriate actions, explicit feedback indicating culturally inappropriate actions, hint feedback or neutral feedback indicating somewhat inappropriate cultural behaviour.<br><br>*Cross-cultural concept knowledge*: A quiz with 18 questions was used to test the participants' understanding of cultural behaviour in the fictitious society. The quiz contained questions about declarative knowledge and transfer scenarios. | Using one-way ANOVAs showed that individuals in the beliefs all condition spent significantly more time in quest 1, had more topics of conversation with the NPCs, and talked to more NPCs than in the other two groups.<br><br>Quest 2 showed a higher percentage of correct feedback for the beliefs all group than for the beliefs at end group, with a mean difference of 0.074. A similar effect with a mean difference of 0.066 was shown for beliefs all compared to the no beliefs group in quest 3.<br><br>A mixed ANOVA was conducted to examine the proportion of culturally-appropriate actions (e.g., correct feedback) across quests and conditions. There was no statistically significant main effect of condition. Additionally, a statistically significant main effect for time was observed for the change in correct feedback across quests. Furthermore, there was a significant interaction effect between the proportion of correct feedback and condition. A closer look showed the same significant mean differences as in the previous one-way ANOVAs.<br><br>There was a significantly higher score in cross-cultural concept knowledge for the beliefs all conditions than for the no beliefs condition (difference in means = 0.61).<br><br>In mixed ANOVAs proportion of correct feedback was associated with number of different parts in communication, attempts to build positive rapport and condition. Cross-cultural concept knowledge was also significantly related to attempts to build positive rapport and condition.<br><br>In a person-centred cluster analysis, four clusters were identified: Fast learning gains (*n* = 119), no learning gains (*n* = 118), slow learning gains (*n* = 110), and inconsistent learning gains (*n* = 86). $\chi^2$-tests revealed that learning cluster was associated with condition, showing that more belief all participants were in the fast learning gain cluster than expected. |

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| | | | | ANCOVAs showed a large significant effect for differences in quiz performance between clusters ($\eta^2 = .28$). participants in the fast learning gains cluster had higher quiz scores than participants in the no learning gains cluster. There was no significant interaction of learner cluster and condition. |
| *Jarrell et al. (2021)* | | | | |
| *N* = 265 male and White participants from the US recruited via MechanicalTurk, the original sample consisted of 691 participants (410 male, 273 female, 7 non-binary) | *Self-developed 2D puzzle game* [Self-developed Games]<br><br>A custom-designed gamefor research, focusing on issues of race and gender representation in video games. The game highlights the lack of diverse protagonists and the prevalence of stereotypes, such as Black characters being portrayed mainly as athletes or gangsters, and female characters as overly sexualized or helpless. Players are prompted to critically reflect on representation | Participants were told that the study was about the relationship between gaming and literacy.<br><br>Participants were randomly assigned to play one level of the game as one of four characters (White male, White female, Black male, Black female). After the game, they were asked about their gaming experience and their identification with the character. Then, they were presented with a pre-fabricated LinkedIn profile of a student interested in a position in game development. The profile randomly differed in photo and name (White male, White female, Black male, Black female). They were asked to rate this profile concerning their likability, technical skills and quality. Finally, participants were asked if they guessed what the study was really about, and then they were debriefed. | *Immersion*: Four items on how much the participants were absorbed in the game (Agarwal & Karahanna, 2000; Saunders et al., 2011).<br><br>*Player Character Identification - Similarity*: Five items on how much the game character suited oneself (Van Looy et al., 2012).<br><br>*Player Character Identification - Embodied Presence*: Five questions on how connected one felt to the game characters (Van Looy et al., 2012).<br><br>*Applicant technic skill*: Three questions on how to assess the applicant's competences in game design (Karmarkar & Tormala, 2010; Ohanian, 1990).<br><br>*Application quality*: Six items on how convincing the | Data was analysed using structural equation models. Further, separate analyses were conducted for participants suspecting the real intent of the study and participants who were unsuspecting.<br><br>All participants rated the technical competence of female applicants as significantly lower than male applicants (0.321 standard deviations). This difference was smaller for the unsuspecting group but remained significant (0.281 standard deviations). Only the unsuspecting group assessed the technical competence of Black applicants significantly lower than White applicants (0.432 standard deviations). There was an additive but not multiplicative effect of bias toward female and Black applicants.<br><br>There was no significant interaction between the ethnicity of the game character and of the applicant. However, a significant effect was found for the interaction between the gender of the character and the applicant. Namely, participants playing a male character estimated the technical skills of female applicants lower than of male applicants (0.281 standard deviations). In contrast, the female character group estimated the technical skills of female applicants higher than for male applicants (0.204 standard deviations). There was no significant interaction of race and gender.<br><br>Higher embodied presence in the Black character group was found to lead to higher ratings of technical ability and likeability for the female Black applicant. For particularly |

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| | patterns and industry practices, linking gameplay to broader discussions about diversity in gaming. | | profile was perceived (X.-P. Chen et al., 2009).<br>*Applicant likeability*: Five items on the perceived personality of the applicant (Srivastava et al., 2009). | high levels of embodied presence, the Black game character condition rated the technical ability (0.988 standard deviations) and likability (1.338 standard deviations) for the female Black applicant higher than for the White male applicant. |
| *Saleem & Anderson (2013)* | | | | |
| Experiment 1: $N$ = 204 undergraduate students from the US (49.02 % female, 80.39 % White, $M_{age}$ = 19.57, $SD$ = 1.71)<br><br>Experiment 2: $N$ = 250 undergraduate students from the US (Before dropout 28.1 % female, $M_{age}$ = 19.42, $SD$ = 1.78) | Experiment 1: *Counter Strike* [FPS]<br><br>A tactical multiplayer FPS where players join either a terrorist or counter-terrorist team and complete objectives such as planting/defusing bombs or rescuing hostages. The game emphasizes fast-paced shooting mechanics, team coordination, and strategy in modern military and terrorism contexts. In the experimental conditions, participants | Experiment 1: Participants were told that the study examined the effects of video games on visual attention. Then, they rated their sources of information about various ethnic groups. Afterwards, participants were randomly assigned to play one of the three games for 30 minutes.<br>After the game, an implicit association test was administered, followed by a questionnaire on demographic information, explicit attitudes and video game evaluation and then completed a drawing task. Finally, they were probed for suspicion before being debriefed about the goals of the study.<br><br>Experiment 2: The procedure was similar to Experiment 1 following a 2 terrorism (present or absent) x 2 Arab character (present or absent) design. Two more game conditions were added, and the | Experiment 1:<br>*Arab information source*: Four items were used to measure whether participants obtained their knowledge of the Arab world from media or from personal sources.<br>*Implicit attitude*: An implicit association test was used to measure attitudes toward Arabs. For this, ten European and ten Arabic names (Park et al., 2007) were juxtaposed with ten pleasant and ten unpleasant words (Greenwald et al., 1998).<br>*Explicit attitudes*: Semantic Differential Items, Attitudes Toward Other Groups (Pratto et al., 1994), and Blatant and Subtle Prejudice (Pettigrew & Meertens, 1995). | Experiment 1: A one-way ANCOVA using Arab information source as covariate showed a significant difference in implicit attitude between the three game conditions. Participants in the Arab-terrorist condition showed significantly greater implicit anti-Arab attitudes than those who played the nonviolent game with a small effect ($d$ = 0.42). The mean of the Russian-terrorist condition did not differ significantly from the other groups. There were small significant effects of the condition on explicit attitudes, with the non-violent condition having significantly lower scores than those with Russian ($d$ = .30) and Arab terrorists ($d$ = .34).<br>For the drawing task, a mixed-model repeated measures 3 (game) x 2 (ethnicity) x 2 (sex) ANCOVA was conducted on stereotypic appearance and affect, with Arab information source as a covariate. The analysis showed a significant main effect for stereotypic appearance. The stereotypic appearance mean for the group with the Arab terrorists was significantly higher with a medium-sized effect compared to both the golf game ($d$ = 0.51) and the one with Russian terrorists ($d$ = 0.51). A significant planned interaction contrast showed that that the game condition only influenced the stereotypic appearance of Arabic painted subjects. Direct comparison between the conditions concerning Arab stereotypic appearances showed medium ($d_{\text{Arab vs. Russian}}$ = 0.50, $d_{Arab\ vs.\ nonviolent}$ = 0.70) to small effects ($d_{Russian\ vs.\ nonviolent}$ = 0.20). |

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| | played matches in which the terrorist team was represented either by Russian or Arabic characters.<br><br>*3D Ultra Mini Golf Adventures* [Arcade-style Games]<br><br>A light-hearted mini-golf simulation featuring themed courses, obstacles, and cartoon-style visuals. Gameplay emphasizes casual, family-friendly fun, focusing on precision and timing.<br><br>Experiment 2: *Arabian Lords* [Strategy Games]<br><br>A real-time strategy and resource management | drawing task was replaced with the measure for *Feelings toward other groups*.<br>To assess the interaction of both factors, Experiment 2 grouped the game conditions in a 2x2 logic, with the first factor being the *absence/presence of a terrorist theme* and the second factor being the *absence/presence of Arab characters*. | *Drawing task*: Participants were given 24 colored pencils to draw "typical" Arab and Caucasian men and women. The four painted persons were coded by two raters on three categories. Affect was judged from the facial expression. Stereotypic appearance was judged based on the number of stereotypic attributes, like blue eyes or hair covering. Lastly, it was coded whether the person is carrying a weapon.<br>Experiment 2:<br>*Arab information source*: Measured in the same way as in Experiment 1.<br>*Implicit Arab attitudes*: Measured in the same way as in Experiment 1.<br>*Explicit Arab attitudes*: Measured in the same way as in Experiment 1.<br>*Feelings toward other groups*: Participants rated their feelings toward Arabs, Latinos, and African Americans using 10 items: positive, negative, anger, fear, furious, afraid, threat, | Since many of the Arab women were painted with their faces covered, only male paintings were analysed for affect. The condition had a significant effect on affect. Participants who played the Arab terrorist game drew characters with significantly less positive affect than those who played the nonviolent game ($d$ = .65, medium effect) or the Russian terrorist game ($d$ = 0.44, small effect). The affect variable also showed that the game condition only had a significant influence on Arab drawings, with large ($d_{\text{Arab vs. nonviolent}}$ = .80) to small effects ($d_{\text{Arab vs. Russian}}$ = 0.48, $d_{\text{Russian vs. nonviolent}}$ = .35).<br>The weapons measure was coded dichotomously, so Chi-square analyses were employed. The proportion of participants who included a weapon in their drawings of Arab males significantly varied depending on the video game they played. The specific ratios for weapons drawn in the Arab-terrorist, Russian-terrorist, and nonviolent game conditions were 8/75 (10.7%), 4/78 (5.1%), and 0/65 (0%).<br><br>Experiment 2: ANCOVAs with Arab information source as a covariate showed a significant small main effect for the presence of terrorism on implicit anti-Arab attitudes ($d$ = .44). Participants who had played a game featuring terrorism reported higher implicit anti-Arab attitudes than those in nonviolent games. There was no significant main effect for the presence of Arab characters and no interaction.<br>Contrast analyses for the comparison of the condition playing *Unreal Tournament*, a game featuring violence but no terrorism, with the two conditions with games that featured violence and terrorism, showed a significant small effect ($d$ = 0.27). Similar, there was a small significant effect for the presence of terrorism and violence ($d$ = 0.49) of two games with Arab characters. There was no |

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| | game. Players take on the role of a merchant striving to expand their business by trading goods, constructing buildings, and competing for influence. The game emphasizes economic growth and cultural setting.<br><br>*Unreal Tournament 2004* [FPS]<br>(see Eastin et al., 2009) | | disgust, pride, and hostility (Mackie et al., 2000; Strahan & Gerbasi, 1972). Only the responses for Arabs were used. | significant difference between Arab game condition without violence and the golf game condition.<br>For explicit Arab attitudes, there were significant small main effects for the presence of terrorism ($d$ = .49) and the presence of Arab characters ($d$ = 0.38). There was no significant interaction between both factors. No significant difference was found between the two non-violent games and the violent non-Arab game. The Arab game without violence had a significantly lower score for explicit anti-Arab attitudes than the Arab game with violence with a small effect ($d$ = .28), but a higher score than the golf game with a small effect ($d$ = .33).<br>Negative affect toward Arabs also showed a significant medium main effect for the presence of terrorists ($d$ = .55) and significant small main effect for the presence of Arab characters ($d$ = .31). There was no significant interaction. Players of the violent non-Arab game showed significantly lower negative attitudes toward Arabs than the players of terrorist games with a small effect ($d$ = .31). There was no significant difference to the nonviolent games. Similarly, the nonviolent Arab game showed significantly lower negative affect toward Arabs with a small effect ($d$ = .46) and no significant difference from the golf game. |
| *Vang & Fox (2014)* | | | | |
| $N$ = 99 White undergraduate students from the US (60.6 % female, $M_{age}$ = 20.57, $SD$ = 3.44) | *Second Life* [Simulation and Social Games]<br><br>An online virtual world where players design and customize avatars, build and modify environments, and | A character resembling the participants' own appearance was prepared. The participants were randomly assigned to one of four conditions, which differed in whether they were paired with a *Black or White digital partner* in the game and whether they were to play a word scramble game *competitively or cooperatively*. | *Partner assessment:* Participants were asked to assess the character of their digital partner in 14 items (Tan et al., 2000).<br>*Agency belief*: People were asked to indicate how convinced they are that the digital partner is controlled by a real person. | Partner assessment was compared between the four groups using ANCOVA, with agency beliefs as a covariate.<br>There was a significant small main effect of partner ethnicity ($\eta^2$ = .04). The Black partner was evaluated more positively than the White partner ($M$ = 2.42, $SD$ = 0.94 to $M$ = 2.64, $SD$ = 0.71). No significant main effect was found for cooperative vs. competitive playing. There was no significant interaction. |

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| | interact with others in shared digital spaces. Gameplay emphasizes social interaction, creativity, and community participation, without predefined goals or endings. | | | |
| *Yang et al. (2014)* | | | | |
| Experiment 1: *N* = 126 White undergraduate students (40 % female)<br><br>Experiment 2: *N* = 141 White undergraduate students (65 % female) | Experiment 1: *Saints Row 2* [Third-person Action Adventure Games] (see Behm-Morawitz, Hoffswell, et al. 2016)<br><br>Experiment 2: *WWE Smackdown vs. RAW* [Sports Game]<br><br>A wrestling simulation game where players control wrestlers in matches featuring varied | Experiment 1: Participants were randomly assigned to one of four groups to play *Saints Row 2* for 20 minutes: 2 ethnicity of the avatar (Black or White) x 2 task (non-violent or violent). The Black avatar differed in visual features and used stereotypical inner-city dialect). After the game, participants completed an implicit association test and a questionnaire.<br><br>Experiment 2: Participants were randomly assigned to one of four groups: 2 game (*WWE Smackdown vs. RAW* or *Fight Night Round 4*) x 2 avatar (Black or White*)*. After the game, participants completed an implicit association test. Then the participants were informed that they would take part in another | Experiment 1:<br>*Explicit negative attitudes toward Black persons*: Symbolic Racism 2000 Scale (Henry & Sears, 2002).<br>*Implicit negative attitudes toward Black persons:* Pictures of Black and White people should be matched to positive or negative words using an implicit association test (IAT).<br><br>Experiment 2:<br>*Implicit attitudes linking Blacks to weapons*: Pictures of Black and White persons were matched to weapons or | Experiment 1: An ANCOVA showed a significant interaction for explicit negative attitudes toward Blacks between the four groups. Participants playing a Black avatar in the condition with violent gameplay had more negative attitudes toward Blacks than the other group with a small effect ($d$ = 0.47). There were no significant main effects for avatar ethnicity or violence.<br>For implicit negative attitudes toward Blacks, an ANCOVA showed no significant effect. Planned contrasts found a significant small effect of playing as a Black avatar with violent gameplay ($d$ = 0.40). Compared to all other conditions, participants playing violently as a Black avatar showed stronger negative attitudes toward the faces of Black persons. This significant small effect was also shown between the two groups with violent gameplay *(*$d$ = 0.37).<br><br>Experiment 2: The Black character group showed significantly higher implicit attitudes linking Blacks to weapons than the White character group with a small effect ($d$ = 0.48). Participants playing a Black avatar furthermore |

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| | fighting styles, moves, and character customization. Gameplay emphasizes physical combat in single-player and multiplayer modes.<br><br>*Fight Night Round 4* [Sports Game]<br>(see Ash, 2016) | study on "food preferences". This was designed to measure aggression. Finally, all participants were informed about the purpose of the study. | harmless objects in an IAT (Nosek et al., 2007).<br>*Aggression*: The Hot Sauce Paradigm was used to measure aggression (Lieberman et al., 1999). | showed significantly higher aggression than White avatar with a medium-sized effect ($d = 0.64$).<br>Further analysis revealed that implicit attitudes mediated the effects of avatar ethnicity on aggression ($b = 0.06$, $p = .008$, 95%BootCI [0.0006, 0.15]). |

**Appendix C**

*Studies Focusing on the Effects of Different Video games: Characteristics of the Sample, the Video game, Study Design, Outcome Variables, Analytical Approach and Main Findings*

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| *Pech & Caspar (2024)* | | | | |
| *N* = 65 male participant from Belgium ($Range_{age}$ = 18 – 36), 1 participant was excluded due to technical errors<br><br>Test game group (*n* = 33, $M_{age}$ = 23.21, *SD* = 4.35);<br>Control game group (*n* = 31, $M_{age}$ = 22.45, *SD* = 2.80) | *Test game: Massive Online Battle Arena (MOBA)* [Self-developed Games]<br><br>A custom-designed Massive Online Battle Arena (MOBA) created for research purposes. Players control an Etanorean character (majority group) in a fictional conflict with the Oldarins (minority group). Through team-based combat and narrative cues, the game highlights injustices against the Oldarins, aiming to implicitly reduce prejudice. | Participants were randomly assigned to play the *test game* or the *control game*. Participants played either game for 2.5 to 4 hours from home. 2 to 4 days after playing, they came to the EEG-lab where their pain threshold for electrical stimulation were assessed. With an installed EEG, they completed 180 trials of the Costly Helping Task and 126 trials of the Intended Prosociality Task. | *Costly Helping Task*: Participants had to repeatedly decide if they wanted to prevent another participant from receiving a mild electric shock at the expense of 0.05€. Participants viewed video recordings of the hands of Black or White fake participants (not) receiving a shock. The number of helping actions toward Black or White participants was used for analysis.<br>*Intended Prosociality Task:* Participants were presented a pair of photos of Black and White persons. They were tasked to decide toward whom they would perform a pre-defined prosocial action. During the task, participants viewed pairs of pictures with different individuals to choose from (free-choice), but also pairs of photos of the same individuals (forced-choice). | For prosocial behaviour, repeated measures ANOVAs with the within-subject factor ingroup vs. outgroup and the between-subject factor game group were conducted for each task. In the costly helping task, the analyses showed a significant small main effect for ingroup vs. outgroup ($\eta^2$ = .051), but not for game group. Participants playing the test game tended to help outgroup individuals significantly more than ingroup individuals with a small effect (*d* = - 0.357). In the intended prosociality task, significantly more intended prosocial actions were aimed toward outgroup individuals with a large effect ($\eta^2$ = .232). There was no significant main effect for game group and no interaction.<br>Similar repeated measures ANOVAs were used to analyse participants' ERPs while witnessing somebody being shocked in the Costly Helping Task. For the P3, there was no significant main effect for ingroup vs. outgroup, game group and no interaction of both. Likewise, pairwise comparisons did not show differences between ingroup and outgroup individuals. There was a significant large main effect for ingroup vs. outgroup for eLPP. Here, participants exhibited a higher amplitude for ingroup individuals. Paired comparisons in revealed that this difference was only apparent for participants playing the control game with a large significant effect (*d* = -0.868). There was no significant main effect for game group and no interaction. The analysis for ILPP showed a similar effect, with a significant large main effect of ingroup vs. outgroup on participant amplitude ($\eta^2$ = .173). Again, paired comparisons showed that a medium-sized effect on amplitude was only present in participants who |

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| | *Control game: A commercial MOBA* [Strategy Games]<br><br>A commercially available MOBA with similar mechanics (e.g., team-based strategic combat and resource management) but without prejudice- or intergroup-related content. The exact title was not disclosed for confidentiality reasons. | | *EEG measures:* Event-related potentials (ERPs) assessed while viewing participants receiving a shock in the Costly Helping task were analysed thrice: at around 300 ms after a stimulus, early, and late Late Positive Potential (i.e., P3, eLPP, ILPP).<br>*Frontal midline theta frequency* (FMθ) power before each keypress in the intended prosociality task was assessed.<br>*Reaction times:* Participants reaction times for decisions in the intended prosociality task were assessed. | played the control game ($d$ = -0.663). There was no main effect for game group and no significant interaction.<br>Concerning FMθ power, there were no significant main effects and no interaction effect. Pairwise comparisons showed that power did not differ between ingroup and outgroup individuals. However, a significant difference in FMθ power with a small effect was detected for participants playing the control game ($d$ = 0.368). In addition, a repeated measures ANOVA with the inner-subject factors pair (outgroup vs. ingroup, ingroup vs. ingroup) and condition (free-choice and forced-choice) and the between-subject factor game group was conducted to assess further differences in FMθ power after pictures were presented. There was a significant interaction between the factors pair and game group, with a medium-sized effect ($\eta^2$ = .071). The authors interpret this as a sign of stronger conflict. The interaction between game group and ingroup vs. outgroup was also significant with a medium-sized effect ($\eta^2$ = .091). A repeated measures ANOVA on participants' reaction times revealed a significant large main effect for condition ($\eta^2$ = .0698) and a medium-sized significant main effect for game group. Reaction times were shorter in the forced-choice condition and longer for participants who had played the test game. There were no other significant main effects or interactions. There were no significant correlations between reaction times or midfrontal theta power and the individual selected in the intended prosociality task. |
| *Peña et al. (2018)* | | | | |
| $N$ = 172 undergraduate students from the US (86.6 % female, $M_{age}$ = 20.15, $SD$ = 2.05, 52.9 % Asian, | *Papers, Please* [Simulation and Social Games]<br><br>A narrative-driven simulation where the player takes | Participants completed the baseline questionnaire prior to the first session. They were randomly assigned to play either *Papers, Please* or *The Westport Independent*. In two sessions on different days but in the same week, | *Intention*: Three items on how much people intend to help immigrants (Novo-Corti et al., 2011).<br>*Attitudes*: Seven items on positive or negative | ANCOVAs were calculated for intention, attitudes, subjective norms, and self-efficacy, respectively, to identify mean differences between the groups and the three measurement time points: before the lab, after the first session, and after the second session. Social dominance was used as a covariate for all analyses. There was no difference |

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| 19.8 % White, 16.3 % Hispanic, 3.5 % African American, 7.6% other)<br><br>*Papers, Please* group ($n$ = 89);<br>*Westport Independent* group ($n$ = 83) | the role of an immigration inspector in a fictional state. Players must decide whether to admit or deny entry to travellers based on increasingly complex immigration rules, balancing efficiency, compassion, and state loyalty.<br><br>*The Westport Independent* [Simulation and Social Games]<br><br>A censorship simulation set in a fictional authoritarian state. Players act as newspaper editors deciding which articles to publish or censor. Choices influence public opinion, government trust, and the | they played one of the two games a total of four times for 45 minutes and completed the questionnaire after each session. | evaluations of immigrants (García et al., 2003).<br>*Subjective norms*: Two items no normative pressures and the approval or disapproval of significant others for helping immigrants (Novo-Corti et al., 2011).<br>*Self-efficacy*: Two items on confidence in the ability to exert control over helping immigrants (Novo-Corti et al., 2011).<br>*Social dominance:* The 16 items in preferences for inequality between social groups were only queried before the lab. | between the two groups for any of the four variables before the experiment.<br>For the variable intention, participants playing *Papers, Please* showed higher expression before the intervention than after the first ($\eta^2$ = .03) and after the second session ($\eta^2$ = .03) with a small effect. There were no significant differences between the second and third measurement time points. Likewise, there were no effects for game condition. For attitudes, there was no main effect or group or time of measurement.<br>For subjective norms, there was no difference between measurement time points and groups. However, a pairwise comparison showed a higher pre-lab score for the *Papers, Please* group than at the post-session scores.<br>The *Papers, Please* group was found to have higher levels of self-efficacy before the intervention compared to after the first session ($\eta^2$ = .11) and second session ($\eta^2$ = .09) with a medium-sized effect. There was no effect between the two sessions. Compared participants playing *The Westport Independent*, participants playing *Papers, Please* showed lower self-efficacy scores to help immigrants after the first ($\eta^2$ = .05) and second sessions ($\eta^2$ = .03) with a small effect. |

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| | newspaper's reputation. | | | |
| *Peña & Hernández Pérez (2020)* | | | | |
| $N$ = 180 Spanish undergraduate students (43.82 % female, $M_{age}$ = 21.41, $SD$ = 10.09)<br><br>*Papers, Please* group ($n$ = 91);<br>*Westport Independent* group ($n$ = 89) | *Papers, Please* [Simulation and Social Games]<br><br>*The Westport Independent* [Simulation and Social Games]<br><br>(see Peña et al., 2018) | This is a replication study of Peña et al. (2018). Therefore, the procedure and variables are analogous. | This is a replication study of Peña et al. (2018). The procedure and variables are analogous.<br>Self-efficacy and subjective norms showed poor reliability and were therefore not used in the analyses. | Repeated measures ANCOVAs were calculated for intention and attitudes respectively, to identify mean differences between the groups and the three measurement time points: before the lab, after the first session, and after the second session. Social dominance was used as a covariate.<br>A priori, participants in the *Paper, Please* group showed a significantly higher score for intention ($\eta^2$ = .037) and attitudes ($\eta^2$ = .098) than the *The Westport Independent* group. The values for attitudes and intentions at baseline were also lower in the current Spanish sample compared to the US American sample from Peña et al. (2018).<br>For Intention, the score in the *Papers, Please* group decreased significantly from baseline to the first ($\eta^2$ = .022) and second session ($\eta^2$ = .087), with small to medium effect sizes. Intention decreased significantly between the first and second session with a large effect ($\eta^2$ = .058). There were no significant differences after the first and second session between the groups.<br>For attitudes, the *Papers, Please* group showed no significant difference between before the lab and after the first session. The score for attitudes was significantly lower after the second session compared to before the lab ($\eta^2$ = .076) and after the first session ($\eta^2$ = .078) with large effects. Compared to *The Westport Independen*t group, attitudes scores remained higher for the *Papers, Please* players after the first ($\eta^2$ = .081) and second sessions ($\eta^2$ = .029) with large effects. |

**Appendix D**

*Studies Focusing on the Effects of Different Formats (Video game versus Text): Characteristics of the Sample, the Video game, Study Design, Outcome Variables, Analytical Approach and Main Findings*

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| (H.-P. Chen et al., 2010) | | | | |
| *N* = 130 sixth-grade elementary students in four classes from Taiwan.<br><br>Experimental group (*n* = 64 from two classes, 45.3 % female);<br>Control group (*n* = 66 from two classes, 45.4 % female) | *FORmosaHope* [Self-developed Games]<br><br>A culture-enriched educational role-playing game designed for 4th–7th graders, integrating science, technology, and society. Players freely explore a virtual Taiwanese village, triggering events that teach about local culture, history, and traditions (e.g., the Tao Tribe). The game also includes 11 educational mini-games . | Over six weeks, all four classes had regular lessons according to the national curriculum guideline. The two classes in the *experimental group* additionally played *FORmosaHope* for one hour per week. The two classes in the *control group* received no extra instruction. Participants answered questionnaires before and after the intervention. | *Questionnaire of Cultural Identities (QCI):* 23 items for the four dimensions:<br>Spiritual Cultural Identities (SpiCI),<br>Material Culture Identities (MtlCI),<br>Scientific Culture Identities (SciCI) and<br>Societal Culture Identities (SocCI). | ANCOVAs were used to compare the means per dimension of the QCI and across the full QCI between pre-test and post-test. A significant main effect of time on the subscales of the QCI and the full QCI was only found in the experimental group. There was also a main effect of condition on the full QCI and each of its subscales. Participants in the experimental group reported significantly higher cultural identity than those in the control group.<br>Also, the score of social identity in the post-test for the experimental group was considered in relation to gender and socioeconomic status. Pre-test scores for the QCI scales were used as covariates. Results indicated that male students' growth in social identity was significantly higher than that of female students. Students with a medium socioeconomic status reported significantly stronger social identity than students with a high socioeconomic status. There was no interaction between gender and socioeconomic status. |
| *Gutierrez et al. (2014)* | | | | |
| *N* = 125 US American undergraduate students (47 % female, 82 % | *FairPlay* [Self-developed Games] | This was an online study. Participants were randomly assigned to *either play the game FairPlay* or to *read a narrative text* about the experiences of Jamal | *Racial bias:* Standardized difference in reaction times between stereotype-congruent trials (White/Black faces with | Only the game group showed a significant correlation between the racial bias score by the implicit association test and empathy (*r* = -.33). There was no overall correlation between racial bias score and empathy. Additional regression analysis revealed that empathy was a significant predictor for |

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| White, 7 % Asian, 4 % underrepresented minority [including Black], 6 % other)<br><br>Game group ($n$ = 58, 43,10% female);<br>Text group ($n$ = 67, 50,75% female) | An educational game where players take the role of Jamal Davis, a Black graduate student navigating life at a predominantly White university. The game emphasizes subtle and unintentional forms of racial discrimination linked to implicit bias (e.g., tokenism, status levelling). When Jamal encounters bias, an in-game Almanac provides just-in-time learning with definitions, examples, and supporting research encouraging players to reflect on the impact of implicit prejudice. | Davis. In both groups, participants were asked to take Jamal's perspective.<br>Subsequently, all participants took an implicit association test. The task was to associate a Black or White face with good or bad terms as quickly as possible. This was followed by a questionnaire. | pleasant/unpleasant words) and stereotype-incongruent trials (White/Black faces with unpleasant/pleasant words) in an Implicit Association Test (IAT).<br>*Bias awareness:* Four items on believes that discriminatory events depicted in the text or game were realistic and whether they could also happen to other minorities.<br>*Perspective taking:* Three items on participants` identification with Jamal Davis in the story or game (only used for manipulation checks).<br>*Empathy:* Six items on participants' feelings of compassion, tenderness, softheartedness, and sympathy (Batson et al., 1997; Lamm et al., 2007). | racial bias in the game group ($B$ = -0.065, $R^2$ = 0.11, $p$ = .014).<br>An ANOVA was used to examine empathy and racial bias between text and game groups. Gender and the interaction of gender per group were analyzed as control variables. The text group had a higher empathy score than the game group ($M$ = 4.33, $SE$ = 0.21 to $M$ = 3.41, $SE$ = .22). A pairwise comparison showed that women ($M$ = 4.89, $SE$ = 0.29) in the text group had a higher score for empathy than men ($M$ = 3.76, $SE$ = 0.30).<br>As hypothesized, it was found that gameplayers with high empathy scores showed significantly lower implicit bias than gameplayers with lower empathy scores: An additional ANOVA examined IAT scores using a median split of the empathy scale as the independent variable. The results indicated that participants in the game group with high empathy ($M$ = 0.137, $SE$ = .076) showed significantly lower racial bias compared to participants with low empathy ($M$ = 0.374, $SE$ = .053). There was no similar effect in the text group. The high empathy game and text groups did not differ significantly. |
| *Lee & Chen (2023)* | | | | |
| $N$ = 156 undergraduate students from the | *Bury Me, My Love* [Role-playing and | Participants were randomly assigned to one of three conditions. In the *Take-Home game condition*, | *Willingness to help refugees:* Likeliness of participants to | A MANOVA showed significant medium-sized effects of condition on empathy and inclusion of others in self (both $\eta^2$ = .08). Tukey HSD post-hoc tests revealed that participants |

| Sample | Video game [Category] | Study design | Outcome variables | Analytical approach and main findings |
|---|---|---|---|---|
| US (75% female, $M_{age}$ = 20.33, *SD* = 1.80, 94 White, 43 Hispanic, 17 Asian, 13 Black, 2 other)<br><br>Take-Home game condition (*n* = 52);<br>In-lab game condition (*n* = 52);<br>News story condition (*n* = 52) | Narrative-driven Games]<br><br>A mobile narrative game simulating the journey of a Syrian refugee couple. Players take the role of Majd, who communicates with his wife via text messages and photos as she travels from Syria to Europe. The story has multiple branching paths and endings depending on the interaction. The game emphasizes emotional engagement and perspective-taking through everyday communication, highlighting the challenges of displacement and migration. | participants played *Bury Me, My love* over the course of one week. They received daily reminders from the researchers to play the game and had to send them a screenshot of the ending they achieved.<br>In the *In-Lab game condition*, participants played the game in one continuous game session in a lab setting. On average, participants completed the game in 40 to 50 minutes.<br>In the *News Story condition*, participants read two featured stories with pictures (length per story approximately 1000 words) about Syrian refugees.<br>All participants completed a pre-test survey and a post-test survey in lab. | help Syrian refugees in the future (Peng et al., 2010).<br>*Empathy toward refugees:* Two subscales on affective empathy and cognitive empathy (Shen, 2010).<br>*Inclusions of others in self:* Participants chose a figure best describing their relationship with a character in the game or news article (adapted from Aron et al., 1992).<br>*Threat perceptions: P*erceived threat from refugees from Middle-Eastern on the subscale realistic threats and symbolic threats.<br>*Intergroup anxiety:* Participants' feeling if they were the only person of their nationality surrounded by a group of strangers from Syria (adapted from (Stephan & Stephan, 1985). | in the in-lab and news conditions reported higher empathy than participants in the take-home condition. For inclusion of others in self, post-hoc tests showed that both game conditions led to less perceived distance to outgroup members than the news condition.<br>The MANOVA also showed a significant small effect of condition on intergroup anxiety ($\eta^2$ = .06). Participants in the take-home game condition reported higher anxiety than those in the news condition. There were no significant differences between the in-lab and news conditions. There was no significant effect of condition on perceived threat.<br>In addition, the authors examined the relationship between the experimental condition and participants' willingness to help refugees. A significant direct effect revealed that participants playing the game at home were less willing to help refugees than those in the news condition (*B* = -.62, *SE* = .27, 95%CI [-1,16, -.08]). There was no similar effect for the in-lab condition. The effects on willingness to help refugees were not mediated through empathy, inclusion of self, intergroup anxiety or perceived threat. |

# 6

# GENERAL DISCUSSION AND CONCLUSION

# 6 General Discussion and Conclusion

Digital games are increasingly shaping how learning environments are designed and how learners engage with both academic content and broader societal issues (e.g., Clark et al., 2016; Costa et al., 2024; Ravyse et al., 2017). While research on DGBL has expanded rapidly, the field continues to face fundamental challenges in establishing clear and interpretable evidence regarding how and why games influence academic outcomes. In addition, evidence on their effects on non-academic outcomes relevant for navigating diverse societies, such as intercultural competence and prejudice, remains even more limited.

This dissertation set out to strengthen the design, interpretability and comparability of evidence in DGBL research by addressing three recurring limitations in the field: weak comparison logic between game and non-game conditions, limited isolation of specific design mechanisms, and fragmentation in outcome constructs and research agendas. To address these challenges, two purposefully designed learning environments (an educational video game and a structurally aligned interactive text application for history learning) were developed and described in two design papers included in this dissertation. These systems enabled controlled comparisons between formats and the systematic manipulation of selected design features.

Building on this foundation, two experimental studies examined how format and design features shape cognitive, motivational, emotional, and behavioural engagement outcomes in the context of digital history learning, with each study targeting specific limitations in prior DGBL research. Given the relative scarcity of research on DGBL in history education (e.g., Clark et al., 2016), the first four papers of this dissertation represent an important step toward a more comprehensive and methodologically rigorous understanding of the potential of digital games in this domain.

This work is complemented by a systematic review examining digital game effects on prejudice and intercultural competence. By extending beyond the domain of history learning, the review addresses fragmentation in outcome constructs and research agendas at the level of the broader DGBL literature. In doing so, it provides insights into the potential of digital games to foster adaptive intercultural outcomes and identifies conceptual and methodological challenges that cut across domains. As will be discussed below, these insights also inform future developments of DGBL in the context of history learning.

Against this background, the aim of the present chapter is to integrate the dissertation’s

findings and to derive conclusions that extend beyond the scope of the individual papers. Section 7.1 synthesizes insights from the development of the *Limes* learning environments. Section 7.2 examines what the experimental studies demonstrate. Section 7.3 systematizes findings on intercultural outcomes. Section 7.4 links the findings from the systematic review to the experimental studies and design work. Section 7.5 discusses next steps for digital history games and intercultural outcomes. Section 7.6 outlines limitations and directions for future research, followed by a final conclusion in Section 7.7.

## 6.1 Insights from the Development of the *Limes* Learning Environments

The development of the educational video game *Limes* and its structurally aligned text-based counterpart provides insights not only into methodological issues in DGBL research but also into the design of digital learning environments for experimental purposes. While these systems served as intervention materials for the empirical studies reported in this dissertation, their development required translating theoretical assumptions about learning processes into concrete design decisions, thereby highlighting key challenges in aligning pedagogical, interactive, and narrative elements across formats.

In particular, the project addressed a recurring methodological challenge in DGBL research: the difficulty of creating experimental conditions that allow meaningful comparisons between game-based and non-game learning formats. Many studies comparing digital games with text-based materials rely on static or passive text conditions that differ substantially from games in terms of interactivity, navigation, and opportunities for behavioural engagement, thereby introducing additional factors that complicate the interpretation of format effects (e.g., Clark et al., 2016; Riopel et al., 2019; Wouters et al., 2013).

In response to this limitation, the present project developed two parallel learning environments: the *Limes* video game and a Unity-based interactive text-app. The text-app replicates key structural features of the game while presenting the same narrative content in text form (see Papers 1 and 2). This required careful design trade-offs to ensure comparability while preserving the integrity of each format, thereby illustrating the complexity of isolating format effects in DGBL research.

Equivalence in instructional content alone is not sufficient when comparing learning formats (e.g., Clark et al., 2016; R. E. Mayer, 2014). Differences in task structure can influence how learners engage with the material and affect learning outcomes, for example through exploration, optional content, or feedback. During the development of the text-app, several gameplay elements from the video game were identified as central for both pedagogical purposes and experimental comparability. These included the optional Codex entries containing additional historical explanations, secondary missions providing optional narrative extensions, time-on-task tracking within different sections of the environment, and a five-star feedback system indicating level completion. Rather than presenting the narrative as a static text, the text-app was designed to reproduce these structures in a non-game format. This design decision

aimed to reduce structural discrepancies between the two learning environments and thereby enable a more valid comparison between game-based and text-based learning conditions.

A second methodological insight relates to the role of behavioural tracking in digital learning environments. Both the video game and the text-app were developed with integrated behavioural logging systems that record learners' interactions with key features of the environment, including time spent on different sections, access to Codex entries, and engagement with optional content. Such behavioural data provide indicators of learners' interaction patterns in digital learning environments, although such measures require careful interpretation and do not directly reflect the quality of learning (e.g., Gašević et al., 2015; Siemens, 2013). The development process therefore illustrates how video games can function not only as instructional tools but also as research instruments capable of generating detailed data on learner behaviour. Integrating behavioural logging during the development stage, as it was done in the present dissertation project, ensures that learning environments can support both pedagogical objectives and empirical investigation of learning processes. Beyond serving as outcome measures that complement test scores and self-reports, behavioural data (e.g., time spent, interaction patterns) provide indicators that help interpret experimental findings, for example by clarifying whether observed differences reflect variation in engagement rather than instructional format alone.

The work on the *Limes* learning environments also highlighted the importance of anticipating the isolation of individual design elements in digital games already in the development process. Digital games typically combine several components at the same time, such as narrative, audiovisual presentation, gameplay mechanics, and interactive choices. Because these elements operate together, it is often unclear which parts of a game influence different types of learning outcomes, including cognitive, motivational, and affective dimensions. This complexity illustrates why carefully designed learning environments are necessary when researchers aim to investigate how specific elements of digital games affect learning outcomes.

A further design-related insight concerns the challenge of balancing player engagement and learning-oriented design during the development of digital learning environments. In the development of *Limes*, this required designing an experience that was both instructive and intrinsically engaging, rather than prioritizing one at the expense of the other. This balance influenced key design decisions. Game elements such as exploration, narrative progression, and interactive choices were incorporated, while the underlying content and task structure were

designed to support learning objectives. Rather than treating “fun” and “learning” as competing goals, the development process aimed to integrate both by embedding learning-relevant content within engaging gameplay activities. This highlights a broader challenge in DGBL: effective learning environments must not only convey instructional content but should also be fun to engage with. Designing for both playful engagement and learning therefore requires careful consideration of how game mechanics, narrative elements, and task structures can simultaneously support enjoyable experiences and meaningful learning processes.

In addition, narrative design (e.g., Naul & Liu, 2020) played an important role in structuring the learning experience. The storyline and characters were developed not only to enhance immersion but also to organize historical information within a coherent context. For example, the player takes on the role of the child of an Alemanni chief, navigating between Germanic and Roman cultural contexts. A central element of the game is the interaction with a Roman centurion, Ariovist, through which learners are exposed to perspectives from both sides. This narrative setup allows historical information to be presented in a balanced way, highlighting differences, commonalities, exchange, and conflict along the frontier. Interactions with characters and story-driven progression provided a framework through which learners could encounter and interpret historical content, thereby supporting engagement with the material without introducing additional instructional differences between conditions.

From a game design perspective, these considerations required translating pedagogical goals into concrete interaction structures. For example, the Codex system was implemented as an optional resource rather than a mandatory component, allowing learners to regulate their engagement with additional information without interrupting the narrative flow. Similarly, historical content was embedded within dialogues and interactions with characters instead of being presented as isolated expository text, ensuring that learning-relevant information remained integrated within the unfolding story. These decisions were not only intended to support engagement but also to avoid introducing additional instructional elements that would differ across experimental conditions. As a result, the design of *Limes* reflects a continuous negotiation between creating an engaging interactive experience, supporting academic outcomes relevant to history learning, and maintaining the level of control required for experimental investigation. While the development of the *Limes* learning environments, from the initial storyline to the final implementation, was time-intensive (approximately two years), it resulted in a self-developed game that provided the necessary flexibility and control to address key research questions at the core of this dissertation.

## 6.2 What Do the Experiments (Jointly) Demonstrate?

Building on the learning environments developed for this project (see Section 7.1), the two experiments examined how different components of the *Limes* system influence learners' motivation, emotions, knowledge acquisition, and behavioural engagement. The first experiment (see Paper 3) compared the educational video game version with a structurally aligned interactive text version presenting the same narrative content. The second experiment (see Paper 4) examined the role of the game's audio layer, including background music, character voices, and narration.

One finding was consistent across experiments: engagement with Codex entries was strongly associated with knowledge acquisition (see also e.g., Schumacher et al., 2025). The Codex system provides optional historical explanations that expand upon the information presented during the main storyline. Analyses using behavioural data showed that participants who accessed these entries more frequently and spent more time reading them achieved higher knowledge scores, even when differences in prior knowledge were controlled for. The consistent association between Codex engagement and knowledge test scores across studies underscores the importance of structural visibility and functional embedding. The principle of intrinsic integration (Habgood & Ainsworth, 2011) argues that learning content is most effective when it is inseparable from core gameplay mechanics. In the present studies, Codex entries were available in all conditions. While they were not mandatory for progression, they were directly embedded in the storyline and game mechanics (or, in the case of the text app, text app mechanics).

On a broader theoretical level, we conceptualized more frequent clicks on codex entries and greater time spent reading as indicators of behavioural engagement, reflecting learners' active interaction with relevant instructional content (e.g., Fredricks et al., 2004; Kiili et al., 2018; von Keyserlingk et al., 2025). Consistent with this perspective, our findings suggest that learners who engaged more intensively with these in-game resources tended to achieve higher knowledge scores, thereby providing support for the role of behavioural engagement in relation to learning processes (Fredricks et al., 2004). However, it is also possible to interpret the findings considering other theoretical frameworks. In the game condition, Codex entries were linked to highlighted green terms in character dialogues that referred to historically relevant concepts, persons, or events. After the dialogue ended, players received an on-screen notification that a new entry had been unlocked, and the Codex menu button displayed a numerical marker until the entry was opened. In the text condition, Codex entries were

implemented as similarly highlighted green clickable links embedded in the written text and corresponded to the same historical content. Research and theory on curiosity (e.g., Murayama et al., 2019) suggest that curiosity arise when individuals become aware of gaps in their knowledge. In the present context, the codex entry links may have made such knowledge gaps salient, thereby prompting information-seeking behaviour. By accessing these links and reading the corresponding entries, participants had the opportunity to reduce these gaps, which likely contributed to subsequent knowledge acquisition.

In the first experiment (Paper 3), a differentiated pattern emerged for motivational and emotional outcomes, underscoring the importance of distinguishing between specific components within outcome domains in DGBL research: Composite measures may obscure meaningful differences between constructs. While participants in the game condition reported higher levels of triggered situational interest and enjoyment, no corresponding differences were observed for maintained interest, and boredom remained low across both conditions.

From the perspective of interest development (Hidi & Renninger, 2006), the pattern found for interest suggests that the game format was effective in capturing learners' initial attention and eliciting immediate interest, probably due to its more dynamic and embodied interaction, such as controlling a character and navigating the environment. However, the absence of effects on maintained interest indicates that these initial motivational responses may not translate into more sustained forms of interest that would support continued engagement with the content over time. This distinction is theoretically important, as triggered and maintained interest are associated with different underlying processes. While triggered interest reflects short-term attentional and affective responses to salient features of the learning environment, maintained interest involves deeper cognitive involvement and a more enduring focus on the content (Hidi & Renninger, 2006; Linnenbrink-Garcia et al., 2010; Renninger et al., 2019). The present findings therefore suggest that game-based features may be particularly effective in initiating interest but may be less effective in supporting its stabilization without additional design elements that promote sustained processing.

Moreover, while it has previously been shown that different types of achievement emotions (e.g., enjoyment, boredom, anxiety) have distinct effects on learning and performances (Pekrun, 2024), the experiment revealed that format effects also differ depending on the academic emotion under investigation. Specifically, higher levels of enjoyment in the game condition indicated that the format was successful in creating a more positively experienced learning situation (Pekrun, 2006; Pekrun & Linnenbrink-Garcia, 2014; Plass et al.,

2015). At the same time, the absence of differences in boredom (combined with generally low levels across conditions) indicates that the comparison condition was not experienced as more disengaging or aversive than the game. Together, these findings show that positive emotional activation and low negative effect may coexist, and that differences between formats may primarily operate through the enhancement of positive experiences rather than the reduction of negative ones. However, it is important to note that the observed effects for both triggered interest and enjoyment were small. One possible explanation is the high quality and structural alignment of the control condition, which included interactive features and opportunities for active engagement, and thus likely provided a more engaging comparison than the static text formats (e.g., PDF-based materials) commonly used in format comparison studies.

In contrast to the motivational and emotional outcomes, the results for knowledge acquisition showed a different pattern in the experiment on format effects (Paper 3). The hypothesis that the game condition would lead to higher knowledge scores had to be rejected, with the text condition instead showing higher performance. This pattern can be interpreted in light of theoretical perspectives emphasizing the role of attentional focus in learning. The FoRe-Squares model (Eitel et al., 2025) proposes that learning outcomes depend, among other factors, on how instructional features guide learners' attention toward task-relevant information and support the processing of that information. In the present study, the interactive and exploratory nature of the game environment may have distributed learners' attention across multiple elements that were not all directly relevant for learning, whereas the text-based format provided a more constrained structure that supported sustained focus on relevant content. This suggests that differences in focus support may explain the observed differences in knowledge outcomes.

Behavioural data provide further insight into these knowledge differences. Participants in the text condition accessed Codex entries more frequently and spent more time engaging with them than participants in the game condition, despite identical availability across formats. As previously discussed, interaction with Codex entries was positively associated with knowledge test performance. This pattern suggests that learners' use of optional, learning-relevant content is linked to differences in knowledge acquisition and may help explain the observed differences between conditions. In addition, these findings illustrate the value of behavioural tracking for refining interpretations of experimental results. By capturing how learners interact with specific content elements, such data provide insights into underlying learning processes that are not accessible through outcome measures alone, thereby enabling a more fine-grained understanding of how different instructional formats shape learning.

In addition to format effects, the first experiment (Paper 3) explored another feature: instructional framing. Instructional framing represents an approach to guiding learners' information processing by directing their attention toward specific goals or aspects of the learning material. In digital learning environments, such framing is often implemented through brief goal instructions prior to task engagement, with the aim of influencing how learners process and prioritize information (e.g., Erhel & Jamet, 2013, 2019). The absence of an effect of instructional framing as well as the absence of interaction effects between format and instructional framing, suggests that minimal goal instructions may have limited impact in complex digital learning environments. Although prior research indicates that goal cues can influence information processing (Erhel & Jamet, 2013, 2019; McCrudden & Schraw, 2007), the present findings imply that such cues may be overshadowed when the structure of the learning environment provides stronger and more continuous guidance for learners' behaviour. This suggests that instructional framing may need to be integrated more directly into the structure of the learning environment, for example through task goals, feedback systems, or progression mechanics, rather than presented only as brief instructions prior to the activity.

Whereas the effects reported in Paper 3 were mixed, with some indications of small format effects on triggered interest and enjoyment, experiment (Paper 4) did not show any significant effect of sound on the outcomes. The absence of significant effects of background sound on knowledge test performance in Study 2 suggests that multimedia enrichment may not automatically enhance learning in complex, exploratory environments. While the cognitive theory of multimedia learning (R. E. Mayer, 2014) proposes that appropriately designed multimodal input can support knowledge construction, applying these principles to video games is not straightforward, as such environments involve interactive and cognitively demanding tasks requiring rapid processing and coordination (e.g., Tobias et al., 2014). In self-paced video games, additional sensory input may not influence learners' allocation of attention to relevant information unless it is functionally integrated into task-relevant processes. The present findings therefore suggest that multimedia elements such as audio may support learning only when they guide attention, provide task-relevant information, or are directly linked to progression mechanics, rather than being added as supplementary features. Similarly, from a FoRe-Squares model perspective (Eitel et al., 2025), stronger effects on knowledge acquisition might be expected if auditory features more directly supported focus-related processes.

Beyond cognitive outcomes, the absence of significant effects of sound on interest, emotions, and behavioural engagement suggests that additional sensory input alone may not

substantially alter how learners engage with or experience the learning environment. Although digital learning environments, and, in particular, specific isolated design features within these environments, can influence multiple outcome domains (see e.g., Loderer et al., 2020, who discuss sound and music as design features that can influence learners' emotional experiences), these effects likely depend on whether design features become functionally relevant for how learners navigate and interpret the material. The absence of significant sound effects in the experiment suggests that the added audio elements did not meaningfully alter how learners directed their attention and interacted with the learning material. Also, looking at behavioural data (i.e., behavioural engagement) reveals that, unlike in the format comparison study, learners' engagement with optional resources remained similar across conditions with and without sound, indicating that the auditory features did not become functionally relevant for guiding exploration or processing. In the present case, the added audio elements may have remained too peripheral to core interaction processes to influence how learners engaged with or processed the content, thereby limiting their impact across the outcome domains examined.

## 6.3 Systematizing Findings on Intercultural Outcomes

The systematic review on digital games and intercultural outcomes was conducted in response to two central issues in the current literature: fragmentation of outcome constructs and fragmentation of research agendas. These issues do not only reflect variability across studies but point to more fundamental inconsistencies in how research on digital games and intercultural outcomes is conceptualized and conducted.

A first key insight concerns the fragmentation of outcome constructs, particularly with regard to intercultural competence. A closer examination of the studies included in the review suggests not only inconsistent operationalization, but a broader disconnect between how intercultural competence is theorized and how it is measured. Although intercultural competence has often been conceptualized as a multidimensional construct encompassing cognitive, motivational, behavioural, and metacognitive components (e.g., Deardorff, 2004; Ang et al., 2007; Van Dyne et al., 2012; Matsumoto & Hwang, 2013), many studies do not provide an explicit conceptual definition of intercultural competence and relied on isolated indicators such empathy or perspective-taking without clarifying how they map onto broader conceptualizations of the construct (e.g., Behm-Morawitz, Pennell, et al., 2016; Gutierrez et al., 2014; Lee & Chen, 2023). In addition, studies including components such as intercultural knowledge or culturally appropriate behaviour are scarce. One notable example is Gjicali et al. (2020), who incorporated both knowledge-related measures and behaviour-oriented elements. This selective focus risks reducing intercultural competence to narrow proxies and may obscure how digital games affect different dimensions of the construct through distinct mechanisms. For example, narrative perspective-taking or embodiment may primarily influence affective or metacognitive dimensions, whereas cooperative intergroup interaction may be more relevant for behavioural dimensions, and information-rich games may contribute more directly to intercultural knowledge. An important implication is that, if outcomes are multidimensional and mechanisms are differentiated, studies relying on single indicators may systematically miss effects occurring in other dimensions. More broadly, the lack of conceptual specification and uneven coverage of theoretically relevant dimensions complicate the interpretation and comparability of findings, hinder cumulative knowledge building, and make it difficult to determine whether studies are capturing the same construct or substantively different outcomes under a shared label.

The review further revealed fragmentation at the level of research agendas. To organize different research strands, the included studies were therefore grouped into several strands that

focus on different mechanisms through which video games are assumed to influence outcomes, and relatedly, on different implementations and research questions. These groups include collaborative gameplay, player-character interaction and embodiment, comparisons between different games with bias-related content, and comparisons between game-based and non-game-based formats. For example, studies on collaborative gameplay are typically based on the idea that intergroup cooperation (see intergroup theory, Allport, 1954) may drive attitudinal change (e.g., Adachi et al., 2015, 2015; Benatov et al., 2021), whereas embodiment-focused studies emphasize perspective-taking processes (e.g., Behm-Morawitz, Hoffswell, et al., 2016; Breves, 2020). Studies examining games with bias-related content focus on how narrative framing and moral context influence intercultural outcomes (e.g., Pech & Caspar, 2024; Peña et al., 2018; Peña & Hernández Pérez, 2020), while format-comparison studies investigate differences across media types (e.g.,H.-P. Chen et al., 2010; Gutierrez et al., 2014; Lee & Chen, 2023).

Findings vary across separate lines of inquiry, as reflected in the patterns observed in the review. First, studies on collaborative gameplay tend to report positive effects on prejudice-related outcomes, such as intergroup attitudes and reduced bias (e.g., Adachi et al., 2015; Benatov et al., 2021). These effects are typically explained by mechanisms of intergroup cooperation. However, the outcomes assessed in these studies are primarily prejudice-focused and there is little evidence of how collaborative gameplay may address components of intercultural competence, such as intercultural knowledge or behavioural skills. Second, research focusing on player-character interaction and embodiment mainly targets processes such as empathy and perspective-taking. Some studies report positive effects on intercultural outcomes (e.g., Behm-Morawitz, Pennell, et al., 2016; Breves, 2020), but findings are not consistent across all interventions. Moreover, these outcomes again reflect specific components rather than a comprehensive assessment of intercultural competence, with little focus on behaviours or knowledge acquisition. Third, studies examining games with bias-related content show mixed results with respect to prejudice-related outcomes (e.g., Pech & Caspar, 2024; Peña et al., 2018). While some interventions report changes in attitudes, others do not find significant effects, suggesting that the effects of bias-related narratives depend on contextual and design factors rather than showing a consistent pattern. Similarly, these studies rarely include measures that would capture broader aspects of intercultural competence. Finally, comparisons between game-based and non-game-based formats do not show a consistent pattern, with effects varying depending on context, outcome measures, and implementation rather than the format alone

(e.g., Gutierrez et al., 2014; Lee & Chen, 2023). For example, differences between conditions appear to depend on specific design and implementation features, such as the form of interaction (e.g., active perspective-taking in gameplay versus passive narrative exposure; Gutierrez et al., 2014) and the implementation context (e.g., controlled in-lab versus longitudinal take-home use; Lee & Chen, 2023), rather than on the format itself.

Taken together, a key contribution of this review is to bring together and synthesize different lines of research on video games and intercultural outcomes. The various designs reflect distinct research questions and underlying assumptions about mechanisms. While this diversity is valuable and the review highlights some consistency within individual lines of research, fragmentation persists, substantial heterogeneity both within and across approaches remains. Nonetheless, by identifying these distinct lines of research following different research agendas, the review may serve as a guiding organizational framework for integrating future findings.

## 6.4 Linking the Findings from the Systematic Review to the Experimental Studies and Design Work

The systematic review presented in Section 6 can be linked to the experimental findings of this dissertation and to the design work. Across the reviewed studies, the effects of digital games on prejudice and intercultural competence vary considerably depending on the interaction structures implemented in the game environment. In particular, several studies reported positive effects when particular gameplay mechanisms were present. For example, collaborative gameplay with members of different social groups has been shown to reduce prejudice and improve intergroup attitudes (Adachi et al., 2015, 2016; Benatov et al., 2021).

This pattern provides a useful context for interpreting the experimental findings reported earlier in this dissertation. As discussed in Section 7.2, the comparison between the *Limes* video game and the structurally aligned interactive text-based version revealed small differences in some motivational and emotional outcomes but not in knowledge acquisition. In addition, isolating the audio layer of the game did not produce significant differences in knowledge, situational interest, academic emotions, or behavioural engagement. Considered together, these results suggest that variation in outcomes may be more strongly related to how learners engage with the content than to the medium or a specific design feature itself. The systematic review in Section 6 indicates that positive effects of video games on prejudice and intercultural outcomes may be most consistent when particular interaction mechanisms are implemented, such as collaborative gameplay or structured intergroup interaction. Even though the experimental manipulations in the present dissertation targeted different outcomes than the studies from the review, it is worth noting that they did not systematically vary such interaction mechanisms. For example, the two formats in Paper 3 differed in presentation, but did not purposefully introduce fundamentally different interaction structures that would be expected to influence learning outcomes (but see the higher engagement with Codex entries in the text condition, which also showed higher knowledge test performance). A similar interpretation applies to the audio manipulation. The absence of effects may indicate that the implemented sound design did not prompt specific interaction patterns that meaningfully altered how learners engaged with the content.

The systematic review also highlights methodological considerations that connect directly to the design work presented in Section 7.1. Many studies included in the review relied on commercially available video games, which can offer important advantages for research. Commercial titles typically provide highly polished interactive environments and familiar

gameplay mechanics that reflect authentic gaming contexts. Examples include the military shooter *Call of Duty: Black Ops*, which was used to examine how cooperative gameplay with outgroup members influences intergroup attitudes (Adachi et al., 2015, 2016), the sandbox game Minecraft, which supported collaborative interaction between Jewish-Israeli and Palestinian-Israeli students (Benatov et al., 2021), the political simulation *PeaceMaker*, in which players assume the role of either the Israeli Prime Minister or the Palestinian President (Alhabash & Wise, 2012, 2014), and the narrative simulation *Papers, Please*, which places players in the role of an immigration officer making decisions about migrants under morally complex conditions (Peña et al., 2018; Peña & Hernández Pérez, 2020).

At the same time, the use of commercial games can limit researchers' ability to manipulate individual design features or align game mechanics closely with specific research questions. In contrast, several studies in the review employed self-developed games designed specifically for experimental purposes, allowing researchers to control narrative content, gameplay structure, and interaction opportunities more precisely. For example, Jarrell et al. (2021) developed a custom *2D puzzle game* to examine how interaction with characters from different social groups influences players' attitudes, while Gutierrez et al. (2014) designed *FairPlay* a point-and-click narrative game that places players in the role of a Black university student navigating subtle racial bias in a university setting.

These examples illustrate how self-developed games can be designed to align specific gameplay mechanics with the research questions under investigation. The development of the *Limes* learning environments, discussed in Section 7.1, follows a similar logic. As described in the design papers, both the educational video game and the structurally aligned text-app were developed to enable controlled comparisons between learning formats while maintaining comparable instructional content and task structures. This design strategy allowed the experimental studies in this dissertation to investigate specific elements of digital learning environments, such as format differences and audiovisual features, while reducing potential confounds arising from differences in narrative material or task structure.

Another insight from the systematic review concerns the subject domains addressed by existing interventions. Many of the reviewed studies focus on contemporary intercultural or social scenarios rather than historical contexts. Within this broader landscape, the *Limes* project represents a complementary approach by applying DGBL methods to history education. The learning environments developed for this dissertation combine narrative storytelling with optional historical explanations and exploratory gameplay elements. This design allows

learners to engage with historical content while also enabling researchers to analyse how different features of digital learning environments influence learning processes.

## 6.5 Next Steps: Digital History Games and Intercultural Outcomes

Even though the *Limes* learning environments developed in the context of this dissertation project focused on academic outcomes, research on digital games for history learning also has the potential to influence intercultural competence and prejudice, which will be reflected on in this section. A potential connection between digital history games and intercultural outcomes follows from the fact that history education engages learners with multiple perspectives and requires them to interpret actions within their historical contexts rather than their own present-day assumptions (Barton & Levstik, 2004; Wineburg, 2001). Intercultural competence involves openness toward other cultures and an interest in engaging with perspectives other than one's own (Byram, 2020). From this perspective, such history learning environments may support components of intercultural competence by fostering openness toward other cultures and facilitating the interpretation of different cultural perspectives.

Historical narratives frequently address encounters, conflicts, and power relations between different social, cultural, or ethnic groups. Engaging with these contexts requires learners to reconstruct how historical actors interpreted their social world and acted within cultural frameworks that may differ substantially from those of the present (Seixas & Morton, 2013; Wineburg, 2001). In this way, historical learning can involve engagement with perspectives that are socially and culturally different from learners' own (Hajisoteriou et al., 2023). In addition, narrative engagement with historically situated characters and events may foster emotional involvement and support processes of historical empathy, enabling learners to engage with the perspectives of historical actors (Kindenberg, 2026). Video games set in historical contexts may foster learners' empathy by enabling interaction with historically situated characters and narratives (Gilbert, 2019). Contextualization may reduce presentism by situating actions within specific historical conditions (Barton & Levstik, 2004; van Drie & van Boxtel, 2008; Wineburg, 2001), which helps counter present-oriented interpretations and fosters more differentiated and context-sensitive understandings of historical actors and social contexts. More generally, multiculturalism climate emphasizes the value of cultural diversity and opportunities to learn about different cultures and their interaction (e.g., Bardach et al., 2024; Rosenthal & Levy, 2010). From this perspective, history as a domain appears particularly well suited to incorporate these aspects into video game environments, for example by

highlighting the value of cultural diversity and providing opportunities to learn about different cultures and their interactions. However, as suggested by the systematic review, such effects likely depend on the extent to which these mechanisms are meaningfully implemented within the learning environment.

## 6.6 Limitations and Directions for Future Research

Despite the insights generated, the present dissertation is subject to limitations that delineate the scope of its conclusions and inform priorities for further investigation.

First, the empirical studies were conducted with university students in a controlled laboratory setting and focused on digital history learning in the context of Roman history. The findings therefore cannot be readily generalized to other populations, age groups, or educational contexts. Replication in school settings and other instructional contexts would help determine whether the observed patterns also hold outside higher education. In addition, the focus on a single historical period constrains the scope of the conclusions. Future research should systematically compare different historical contexts to examine the extent to which observed effects depend on the specific characteristics of the historical content, and their interactions with learner characteristics. In addition, participants had, on average, limited prior knowledge of Roman history. This may have influenced both their engagement with the materials and the observed knowledge gains (e.g., by shaping how they process and integrate information). As a result, the findings may differ for learners with higher levels of prior knowledge.

Second, behavioural log data were systematically collected and indicated a relationship between Codex engagement and knowledge outcomes. However, this association was examined at an aggregate level (e.g., total time spent on Codex entries) and does not capture moment-to-moment engagement with specific content. In addition, outcome measures were assessed at a single time point, limiting conclusions about temporal dynamics and retention. Future research should therefore adopt more fine-grained approaches by linking behavioural engagement at the momentary level to motivational and emotional states. Such designs would allow for a more precise analysis of how interactions with specific game elements relate to motivational and emotional processes as they unfold. Moreover, including delayed knowledge assessments would make it possible to examine whether different formats (e.g., game-based vs. text-based environments) differentially support longer-term retention.

Third, this dissertation investigated just one of many potentially interesting design features (e.g., Plass et al., 2015). Moreover, the sound manipulation (Paper 4) targeted ambient background audio rather than functionally instructional auditory elements such as spoken explanations or signalling cues. While no effects were observed for background sound, the manipulation was limited to ambient audio and therefore does not exclude the possibility that other forms of auditory integration may influence attentional allocation and cognitive

processing. Future investigations should differentiate between ambient, narrative, and instructional audio to clarify how auditory features interact with task relevance.

Fourth, although learner characteristics (e.g., prior knowledge) were statistically controlled, they were not a primary focus of the analyses. As a result, the present studies do not provide insight into how individual differences shape engagement with the learning environments. Future research should therefore examine how learners' motivational, emotional and personality characteristics relate to their in-game behaviours and learning outcomes.

Finally, the systematic review showed that key constructs were often measured in inconsistent and narrow ways. While the empirical studies in this dissertation focused on academic outcomes, future research should apply similar experimental designs to assess prejudice and multiple dimensions of intercultural competence outcomes (Matsumoto & Hwang, 2013) outcomes while using clearer and more consistent measurement approaches. Improving construct clarity is essential for cumulative progress in the field.

## 6.7 Conclusion

The present dissertation set out to strengthen the interpretability and comparability of findings in DGBL. By bringing together a range of different areas relevant to DGBL (design work, experimental comparisons, and field-level synthesis), the thesis addressed persistent substantive and methodological challenges in the field, using digital history learning as its primary empirical context. Methodologically, the research advances current knowledge on DGBL by employing structurally aligned control conditions, isolating discrete design elements, and integrating behavioural measurement. A central theoretical implication is that the effects of digital games depend on how content is integrated into the environment and how individuals engage with it. For example, the experimental studies show that behavioural engagement with content-rich historical explanations predicts knowledge acquisition across formats. Similarly, the systematic review demonstrates that intercultural outcomes often depend on specific interactive mechanisms (e.g., cooperation, embodiment, moral framing). Overall, the dissertation contributes to a more coherent and differentiated understanding of DGBL.

# 7

# REFERENCES

# 7 References

Adachi, P., Hodson, G., Willoughby, T., Blank, C., & Ha, A. (2016). From Outgroups to Allied Forces: Effect of Intergroup Cooperation in Violent and Nonviolent Video Games on Boosting Favorable Outgroup Attitudes. *Journal of Experimental Psychology: General*, *145*(3), 259–265. APA PsycInfo (WOS:000370623400001). https://doi.org/10.1037/xge0000145

Adachi, P., Hodson, G., Willoughby, T., & Zanette, S. (2015). Brothers and Sisters in Arms: Intergroup Cooperation in a Violent Shooter Game Can Reduce Intergroup Bias. *Psychology of Violence, Protective Factors, Resilience, and Violence*, *5*(4), 455–462. APA PsycInfo (WOS:000362511800014). https://doi.org/10.1037/a0037407

ADL. (2004). *Attitudes Toward Jews, Israel and the Palestinian-Israeli Conflict in Ten European Countries* [Report]. Anti-Defamation League. (New York, NY.). https://archive.jpr.org.uk/object-2740

Agarwal, R., & Karahanna, E. (2000). Time Flies When You're Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage. *MIS Quarterly*, *24*(4), 665–694. https://doi.org/10.2307/3250951

Alhabash, S., & Wise, K. (2012). PeaceMaker: Changing Students' Attitudes Toward Palestinians and Israelis Through Video Game Play. *International Journal of Communication*, *6*, 356–380. (WOS:000302535500003).

Alhabash, S., & Wise, K. (2014). Playing their game: Changing stereotypes of Palestinians and Israelis through videogame play. *New Media & Society*, *17*(8), 1358–1376. APA PsycInfo (WOS:000359263100009). https://doi.org/10.1177/1461444814525010

All, A., Castellar, E. P. N., & Looy, J. V. (2014). Measuring Effectiveness in Digital Game-Based Learning: A Methodological Review. *International Journal of Serious Games*, *1*(2), Article 2. https://doi.org/10.17083/ijsg.v1i2.18

Allport, G. W. (1954). *The nature of prejudice* (pp. xviii, 537). Addison-Wesley.

Altemeyer, B. (1988). *Enemies of freedom: Understanding right-wing authoritarianism* (1st

ed). Jossey-Bass Publishers.

Anderson, C. A. (1999). *Coding key for the word completion task*. Institution: Iowa State University, Department of Psychology.

Anderson, C. A., Carnagey, N. L., & Eubanks, J. (2003). Exposure to violent media: The effects of songs with violent lyrics on aggressive thoughts and feelings. *Journal of Personality and Social Psychology*, *84*(5), 960–971. https://doi.org/10.1037/0022-3514.84.5.960

Anderson, C. A., Carnagey, N. L., Flanagan, M., Benjamin Jr., A. J., Eubanks, J., & Valentine, J. C. (2004). Violent Video Games: Specific Effects of Violent Content on Aggressive Thoughts and Behavior. In *Advances in experimental social psychology, Vol. 36* (pp. 199–249). Elsevier Academic Press. https://doi.org/10.1016/S0065-2601(04)36004-1

Anderson, C. A., Deuser, W. E., & DeNeve, K. M. (1995). Hot Temperatures, Hostile Affect, Hostile Cognition, and Arousal: Tests of a General Model of Affective Aggression. *Personality and Social Psychology Bulletin*, *21*(5), 434–448. https://doi.org/10.1177/0146167295215002

Appiah, O. (2001). Ethnic identification on adolescents' evaluations of advertisements. *Journal of Advertising Research*, *41*(5), 7–22. https://doi.org/10.2501/JAR-41-5-7-22

Appiah, O. (2006). Rich Media, Poor Media: The Impact of Audio/Video vs. Text/Picture Testimonial Ads on Browsers'Evaluations of Commercial Web Sites and Online Products. *Journal of Current Issues & Research in Advertising*, *28*(1), 73–86. https://doi.org/10.1080/10641734.2006.10505192

Appiah, O. (2007). The Effectiveness of "Typical-User" Testimonial Advertisements on Black and White Browsers' Evaluations of Products on Commercial Websites: Do They Really Work? *Journal of Advertising Research*, *47*(1), 14–27. https://doi.org/10.2501/S0021849907070031

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, *63*(4), 596–612. https://doi.org/10.1037/0022-3514.63.4.596

Ash, E. (2016). Priming or Proteus Effect? Examining the Effects of Avatar Race on In-Game Behavior and Post-Play Aggressive Cognition and Affect in Video Games. *Games and Culture: A Journal of Interactive Media*, *11*(4), 422–440. APA PsycInfo (WOS:000375769800004). https://doi.org/10.1177/1555412014568870

Bachen, C., Hernández-Ramos, P., & Raphael, C. (2012). Simulating REAL LIVES: Promoting global empathy and interest in learning through simulation games. *Communication*. https://scholarcommons.scu.edu/comm/32

Bardach, L., Röhl, S., Oczlon, S., Schumacher, A., Lüftenegger, M., Lavelle-Hill, R., Schwarzenthal, M., & Zitzmann, S. (2024). Cultural diversity climate in school: A meta-analytic review of its relationships with intergroup, academic, and socioemotional outcomes. *Psychological Bulletin*, *150*(12), 1397–1439. https://doi.org/10.1037/bul0000454

Bar-Tal, D., & Labin, D. (2001). The effect of a major event on stereotyping: Terrorist attacks in Israel and Israeli adolescents' perceptions of Palestinians, Jordanians and Arabs. *European Journal of Social Psychology*, *31*(3), 265–280. https://doi.org/10.1002/ejsp.43

Barton, K. C., & Levstik, L. S. (2004). *Teaching History for the Common Good* (0 ed.). Routledge. https://doi.org/10.4324/9781410610508

Batson, C. D., Early, S., & Salvarani, G. (1997). Perspective Taking: Imagining How Another Feels Versus Imaging How You Would Feel. *Personality and Social Psychology Bulletin*, *23*(7), 751–758. https://doi.org/10.1177/0146167297237008

Bayerisches Staatsministerium für Unterricht und Kultus. (2026). *LehrplanPLUS Gymnasium: Geschichte, Jahrgangsstufe 6*. LehrplanPLUS Bayern. https://www.lehrplanplus.bayern.de/fachlehrplan/gymnasium/6/geschichte

Behm-Morawitz, E. (2013). Mirrored selves: The influence of self-presence in a virtual world on health, appearance, and well-being. *Computers in Human Behavior, Including Special Section Youth, Internet, and Wellbeing*, *29*(1), 119–128. https://doi.org/10.1016/j.chb.2012.07.023

Behm-Morawitz, E., Hoffswell, J., & Chen, S. (2016). The Virtual Threat Effect: A Test of Competing Explanations for the Effects of Racial Stereotyping in Video Games on Players' Cognitions. *Cyberpsychology, Behavior, and Social Networking*, *19*(5), 308–313. APA PsycInfo (WOS:000376685100003). https://doi.org/10.1089/cyber.2015.0461

Behm-Morawitz, E., Pennell, H., & Speno, A. (2016). The effects of virtual racial embodiment in a gaming app on reducing prejudice. *Communication Monographs*, *83*(3), 396–418. APA PsycInfo (WOS:000382574900006). https://doi.org/10.1080/03637751.2015.1128556

Benatov, J., Berger, R., & Tadmor, C. T. (2021). Gaming for peace: Virtual contact through cooperative video gaming increases children's intergroup tolerance in the context of the Israeli–Palestinian conflict. *Journal of Experimental Social Psychology*, *92*, 104065. https://doi.org/10.1016/j.jesp.2020.104065

Berger, R., Brenick, A., Lawrence, S. E., Coco, L., & Abu-Raiya, H. (2018). Comparing the effectiveness and durability of contact- and skills-based prejudice reduction approaches. *Journal of Applied Developmental Psychology*, *59*, 46–53. https://doi.org/10.1016/j.appdev.2018.04.002

Boeker, M., Andel, P., Vach, W., & Frankenschmidt, A. (2013). Game-Based E-Learning Is More Effective than a Conventional Instructional Method: A Randomized Controlled Trial with Third-Year Medical Students. *PLOS ONE*, *8*(12), e82328. https://doi.org/10.1371/journal.pone.0082328

Boyle, E. A., Connolly, T. M., Hainey, T., & Boyle, J. M. (2012). Engagement in digital entertainment games: A systematic review. *Computers in Human Behavior*, *28*(3), 771–780. https://doi.org/10.1016/j.chb.2011.11.020

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., Ribeiro, C., & Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education*, *94*, 178–192. https://doi.org/10.1016/j.compedu.2015.11.003

Boyle, E., Hauge, J. B., Leith, M., Sim, D., Hummel, H., Jandric, P., & Jimoyiannis, A. (2019).

*Linking Learning Outcomes and Game Mechanics in the Early Stages of the RU EU? Project* (M. Gentile, M. Allegra, & H. Sobke, Eds.; Vol. 11385, pp. 191–200). (WOS:000612997900018). https://doi.org/10.1007/978-3-030-11548-7_18

Breves, P. (2020). Reducing Outgroup Bias through Intergroup Contact with Non-Playable Video Game Characters in VR. *Presence-Virtual and Augmented Reality*, *27*(3), 257–273. (WOS:000553161100001). https://doi.org/10.1162/pres_a_00330

Byram, M. (2020). *Teaching and Assessing Intercultural Communicative Competence: Revisited*. Multilingual Matters.

Byun, J., & Loh, C. S. (2015). Audial engagement: Effects of game sound on learner engagement in digital game-based learning environments. *Computers in Human Behavior*, *46*, 129–138. https://doi.org/10.1016/j.chb.2014.12.052

Carenys, J., & Moya, S. (2016). Digital game-based learning in accounting and business education. *Accounting Education*, *25*(6), 598–651. https://doi.org/10.1080/09639284.2016.1241951

Çelik, T. (2023). Using Digital Games to Teach History: A Design-Based Study. *Research on Education and Media*, *15*(2), 10–24. https://doi.org/10.2478/rem-2023-0020

Chapman, A. (2016). *Digital Games as History: Adam Chapman*. https://www.book2look.com/book/DdgRKX4PQ7

Chen, H.-P., Lien, C.-J., Annetta, L., & Lu, Y.-L. (2010). The Influence of an Educational Computer Game on Children's Cultural Identities. *Educational Technology & Society*, *13*(1), 94–105. eric (International Forum of Educational Technology & Society. Athabasca University, School of Computing & Information Systems, 1 University Drive, Athabasca, AB T9S 3A3, Canada. Tel: 780-675-6812; Fax: 780-675-6973; Web site: http://www.ifets.info).

Chen, V., Koek, W., & Bowman, N. (2025). The effect of in-game moral choices and NPCs' identities on players' intergroup attitudes. *Entertainment Computing*, *52*. (WOS:001259668500001). https://doi.org/10.1016/j.entcom.2024.100766

Chen, X.-P., Yao, X., & Kotha, S. (2009). Entrepreneur Passion And Preparedness In Business Plan Presentations: A Persuasion Analysis Of Venture Capitalists' Funding Decisions. *Academy of Management Journal*, *52*(1), 199–214. https://doi.org/10.5465/amj.2009.36462018

Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital Games, Design, and Learning: A Systematic Review and Meta-Analysis. *Review of Educational Research*, *86*(1), 79–122. https://doi.org/10.3102/0034654315582065

Correll, J., Park, B., Judd, C., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, *83*(6), 1314–1329. (WOS:000179661800004). https://doi.org/10.1037//0022-3514.83.6.1314

Correll, J., Park, B., Judd, C., & Wittenbrink, B. (2007). The influence of stereotypes on decisions to shoot. *European Journal of Social Psychology*, *37*(6), 1102–1117. APA PsycInfo (WOS:000251387600002). https://doi.org/10.1002/ejsp.450

Costa, L. P., Bierwiaczonek, K., & Bianchi, M. (2024). Does Digital Intergroup Contact Reduce Prejudice? A Meta-Analysis. *Cyberpsychology, Behavior, and Social Networking*, *27*(7), 440–451. https://doi.org/10.1089/cyber.2023.0591

Crivellari, C. (2010). *Analysis of National Curricula of Geography and History*. https://repositorio.grial.eu/server/api/core/bitstreams/774dd48c-5132-49a3-9d56-f0471f4680b8/content

Deardorff, D. K. (2006). Identification and Assessment of Intercultural Competence as a Student Outcome of Internationalization. *Journal of Studies in International Education*, *10*(3), 241–266. https://doi.org/10.1177/1028315306287002

Deci, E. L., & Ryan, R. M. (2012). Motivation, personality, and development within embedded social contexts: An overview of self-determination theory. In *The Oxford handbook of*

*human motivation* (pp. 85–107). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195399820.013.0006

Deterding, S. (2012). Gamification: Designing for motivation. *Interactions*, *19*(4), 14–17. https://doi.org/10.1145/2212877.2212883

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification." *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek '11*, 9–15. https://doi.org/10.1145/2181037.2181040

Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, *23*(3), 316–326. https://doi.org/10.1177/0146167297233009

Eastin, M. (2006). Video Game Violence and the Female Game Player: Self- and Opponent Gender Effects on Presence and Aggressive Thoughts. *Human Communication Research*, *32*(3), 351–372. https://doi.org/10.1111/j.1468-2958.2006.00279.x

Eastin, M., Appiah, O., & Cicchirllo, V. (2009). Identification and the Influence of Cultural Stereotyping on Postvideogame Play Hostility. *Human Communication Research*, *35*(3), 337-U10. APA PsycInfo (WOS:000266706000002). https://doi.org/10.1111/j.1468-2958.2009.01354.x

Eccles, J. S., & Wigfield, A. (2002). Motivational Beliefs, Values, and Goals. *Annual Review of Psychology*, *53*(Volume 53, 2002), 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, *61*, 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

Eitel, A., Krebs, M.-C., & Schöne, C. (2025). Introducing the FoRe-Squares Model: Focus Support and Reward as Key Predictors of Effective Technology-Augmented Instruction. *Educational Psychology Review*, *37*(3), 67. https://doi.org/10.1007/s10648-

025-10045-z

Elliot, A. J. (1999). Approach and avoidance motivation and achievement goals. *Educational Psychologist*, *34*(3), 169–189. https://doi.org/10.1207/s15326985ep3403_3

Erhel, S., & Jamet, E. (2013). Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & Education*, *67*, 156–167. https://doi.org/10.1016/j.compedu.2013.02.019

Erhel, S., & Jamet, E. (2019). Improving instructions in educational computer games: Exploring the relations between goal specificity, flow experience and learning outcomes. *Computers in Human Behavior*, *91*, 106–114. https://doi.org/10.1016/j.chb.2018.09.020

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Gaertner, S. L., & Dovidio, J. F. (2013). *Reducing intergroup bias: The common ingroup identity model*. Psychology.

García, M. C. G., Navas, M. S., Guirado, M. I. C., & Alonso, F. M. (2003). Immigration and prejudice: Attitudes of sample of adolescents from Almeria [in Spanish]. *Acción Psicológica*, *2*(2), 137–147. https://doi.org/10.5944/ap.2.2.527

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Gee, J. P. (2007). *What Video Games Have to Teach Us About Learning and Literacy. Second Edition: Revised and Updated Edition* (2nd ed.). St. Martin's Griffin.

Gilbert, L. (2019). "Assassin's Creed reminds us that history is human experience": Students' senses of empathy while playing a narrative video game. *Theory & Research in Social Education*, *47*(1), 108–137. https://doi.org/10.1080/00933104.2018.1560713

Gjicali, K., Finn, B. M., & Hebert, D. (2020). Effects of belief generation on social exploration, culturally-appropriate actions, and cross-cultural concept learning in a game-based

social simulation. *Computers & Education*, *156*. (WOS:000557983800013). https://doi.org/10.1016/j.compedu.2020.103959

Graafstal, E. P., Willems, W. J. H., & Bödecker, S. (2021). *The lower german limes*. Archaeopress Publishing Ltd.

Graham, J., Haidt, J., & Nosek, B. (2008). *The Moral Foundations Questionnaire*. Moral Foundations. https://moralfoundations.org/questionnaires/

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. https://doi.org/10.1037/0022-3514.85.2.197

Greitemeyer, T. (2013). Playing Violent Video Games Increases Intergroup Bias. *Personality and Social Psychology Bulletin*, *40*(1), 70–78. APA PsycInfo (WOS:000328598200006). https://doi.org/10.1177/0146167213505872

Gutierrez, B., Kaatz, A., Chu, S., Ramirez, D., Samson-Samuel, C., & Carnes, M. (2014). "Fair Play'': A Videogame Designed to Address Implicit Race Bias Through Active Perspective Taking. *Games for Health Journal*, *3*(6), 371–378. (WOS:000363385400006). https://doi.org/10.1089/g4h.2013.0071

Habgood, M. P. J., & Ainsworth, S. E. (2011). Motivating Children to Learn Effectively: Exploring the Value of Intrinsic Integration in Educational Games. *Journal of the Learning Sciences*, *20*(2), 169–206. https://doi.org/10.1080/10508406.2010.508029

Hajisoteriou, C., Solomou, E. A., & Antoniou, M. (2023). What kind of history teaching may promote intercultural competence in culturally-diverse societies? Teachers' reflections from a case of conflict. *Frontiers in Education*, *8*. https://doi.org/10.3389/feduc.2023.1156518

Halem, N. van, Klaveren, C. van, Drachsler, H., Schmitz, M., & Cornelisz, I. (2020). Tracking Patterns in Self-Regulated Learning Using Students' Self-Reports and Online Trace Data. *Frontline Learning Research*, *8*(3), 140–163. https://doi.org/10.14786/flr.v8i3.497

Hartmann, T., Toz, E., & Brandon, M. (2010). Just a Game? Unjustified Virtual Violence Produces Guilt in Empathetic Players. *Media Psychology*, *13*(4), 339–363. https://doi.org/10.1080/15213269.2010.524912

Hartmann, T., Wirth, W., Schramm, H., Klimmt, C., Vorderer, P., Gysbers, A., Böcking, S., Ravaja, N., Laarni, J., Saari, T. A., Gouveia, F. R., & Sacau, A. M. (2016). The spatial presence experience scale (SPES): A short self-report measure for diverse media settings. *Journal of Media Psychology*, *28*(1), 1–15. https://doi.org/10.1027/1864-1105/a000137

Henry, P. J., & Sears, D. O. (2002). The Symbolic Racism 2000 Scale. *Political Psychology*, *23*(2), 253–283. https://doi.org/10.1111/0162-895X.00281

Hessisches Kultusministerium. (2002). *Geschichte: Bildungsgang Hauptschule (Jahrgangsstufen 5 bis 9/10)*. https://kultus.hessen.de/sites/kultus.hessen.de/files/2021-06/lphauptgeschichte.pdf

Hidi, S., & Renninger, K. A. (2006). The Four-Phase Model of Interest Development. *Educational Psychologist*, *41*(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

Hodson, G., Dovidio, J. F., & Esses, V. M. (2003). Ingroup identification as a moderator of positive–negative asymmetry in social discrimination. *European Journal of Social Psychology*, *33*(2), 215–233. https://doi.org/10.1002/ejsp.141

HolmesIV, O., Jiang, K., Avery, D. R., McKay, P. F., Oh, I.-S., & Tillman, C. J. (2021). A Meta-Analysis Integrating 25 Years of Diversity Climate Research. *Journal of Management*, *47*(6), 1357–1382. https://doi.org/10.1177/0149206320934547

Huizenga, J., Admiraal, W., Akkerman, S., & ten Dam, G. (2009). Mobile game-based learning in secondary education: Engagement, motivation and learning in a mobile city game.

*Journal of Computer Assisted Learning*, *25*(4), 332–344. https://doi.org/10.1111/jca.2009.25.issue-410.1111/j.1365-2729.2009.00316.x

Hung, C.-M., Huang, I., & Hwang, G.-J. (2014). Effects of digital game-based learning on students' self-efficacy, motivation, anxiety, and achievements in learning mathematics. *Journal of Computers in Education*, *1*(2), 151–166. https://doi.org/10.1007/s40692-014-0008-8

Jackson, G. T., & McNamara, D. S. (2013). Motivation and performance in a game-based intelligent tutoring system. *Journal of Educational Psychology*, *105*(4), 1036–1049. https://doi.org/10.1037/a0032580

Jarrell, M. A., Anaraky, R. G., Knijnenburg, B. P., & Ash, E. M. (2021). *Using Intersectional Representation & Embodied Identification in Standard Video Game Play to Reduce Societal Biases*. (WOS:000758168001061). CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3411764.3445161

Kaminsky, M., & Bar-Tal, D. (1996). Stereotypic perceptions of different Israeli-Arab labels as a function of age and religiosity. *Studies in Education*, *1*, 121–157.

Kao, D., Ratan, R., Mousas, C., & Magana, A. J. (2021). The Effects of a Self-Similar Avatar Voice in Educational Games. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CHI PLAY), 1–28. https://doi.org/10.1145/3474665

Karmarkar, U. R., & Tormala, Z. L. (2010). Believe Me, I Have No Idea What I'm Talking About: The Effects of Source Certainty on Consumer Involvement and Persuasion. *Journal of Consumer Research*, *36*(6), 1033–1049. https://doi.org/10.1086/648381

Karsenti, T., & Parent, S. (2020). Teaching history with the video game Assassin's Creed: Effective teaching practices and reported learning. *Review of Science, Mathematics and ICT Education*, *14*(1), 27–45. https://doi.org/10.26220/rev.3278

Kawakami, K., & Dovidio, J. F. (2001). The reliability of implicit stereotyping. *Personality and Social Psychology Bulletin*, *27*(2), 212–225. https://doi.org/10.1177/0146167201272007

Kiili, K., Moeller, K., & Ninaus, M. (2018). Evaluating the effectiveness of a game-based rational number training—In-game metrics as learning indicators. *Computers & Education*, *120*, 13–28. https://doi.org/10.1016/j.compedu.2018.01.012

Kindenberg, B. (2026). Storytelling in history classrooms: Moving through historical empathy and analysis. *Classroom Discourse*, *0*(0), 1–19. https://doi.org/10.1080/19463014.2025.2598212

Kuran, M. Ş., Tozoğlu, A. E., & Tavernari, C. (2018). History-Themed Games in History Education: Experiences on a Blended World History Course. *2018 17th International Conference on Information Technology Based Higher Education and Training (ITHET)*, 1–8. https://doi.org/10.1109/ITHET.2018.8424767

Lai, C.-H., & Hu, P.-Y. (2025). The Gaming Revolution in History Education: The Practice and Challenges of Integrating Game-Based Learning into Formal Education. *Information*, *16*(6), 490. https://doi.org/10.3390/info16060490

Lamm, C., Batson, C. D., & Decety, J. (2007). The neural substrate of human empathy: Effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience*, *19*(1), 42–58. https://doi.org/10.1162/jocn.2007.19.1.42

Lee, Y., & Chen, M. (2023). Intergroup Contact With a Virtual Refugee: Reducing Prejudice Through a Cooperative Game. *Psychology of Popular Media*, *12*(4), 383–392. APA PsycInfo (WOS:000847817400001). https://doi.org/10.1037/ppm0000433

Lei, H., Cui, Y., & Zhou, W. (2018). Relationships between student engagement and academic achievement: A meta-analysis. *Social Behavior and Personality: An International Journal*, *46*(3), 517–528. https://doi.org/10.2224/sbp.7054

Lieberman, J. D., Solomon, S., Greenberg, J., & McGregor, H. A. (1999). A hot new way to measure aggression: Hot sauce allocation. *Aggressive Behavior*, *25*(5), 331–348. https://doi.org/10.1002/(SICI)1098-2337(1999)25:5%3C331::AID-AB2%3E3.0.CO;2-1

Linnenbrink-Garcia, L., Durik, A. M., Conley, A. M., Barron, K. E., Tauer, J. M., Karabenick, S. A., & Harackiewicz, J. M. (2010). Measuring Situational Interest in Academic

Domains. *Educational and Psychological Measurement*, *70*(4), 647–671. https://doi.org/10.1177/0013164409355699

Linnenbrink-Garcia, L., Patall, E. A., & Pekrun, R. (2016). Adaptive Motivation and Emotion in Education: Research and Principles for Instructional Design. *Policy Insights from the Behavioral and Brain Sciences*, *3*(2), 228–236. https://doi.org/10.1177/2372732216644450

Lipscomb, S. D., & Zehnder, S. M. (2004). Immersion in the Virtual Environment: The Effect of a Musical Score on the Video Gaming Experience. *Journal of PHYSIOLOGICAL ANTHROPOLOGY and Applied Human Science*, *23*(6), 337–343. https://doi.org/10.2114/jpa.23.337

Loderer, K., Pekrun, R., & Plass, J. L. (2020). Emotional foundations of game-based learning. In *Handbook of game-based learning* (pp. 111–151). The MIT Press.

Louw, J., Muller, J., & Tredoux, C. (2008). Time-on-task, technology and mathematics achievement. *Evaluation and Program Planning*, *31*(1), 41–50. https://doi.org/10.1016/j.evalprogplan.2007.11.001

Mackie, D. M., Devos, T., & Smith, E. R. (2000). Intergroup emotions: Explaining offensive action tendencies in an intergroup context. *Journal of Personality and Social Psychology*, *79*(4), 602–616. https://doi.org/10.1037/0022-3514.79.4.602

Mao, P., Cai, Z., Wang, Z., Hao, X., Fan, X., & Sun, X. (2024). The effects of dynamic and static feedback under tasks with different difficulty levels in digital game-based learning. *The Internet and Higher Education*, *60*, 100923. https://doi.org/10.1016/j.iheduc.2023.100923

Marsh, H. W., & Craven, R. G. (2006). Reciprocal Effects of Self-Concept and Performance From a Multidimensional Perspective: Beyond Seductive Pleasure and Unidimensional Perspectives. *Perspectives on Psychological Science*, *1*(2), 133–163. https://doi.org/10.1111/j.1745-6916.2006.00010.x

Martin, F., & Borup, J. (2022). Online learner engagement: Conceptual definitions, research themes, and supportive practices. *Educational Psychologist*, *57*(3), 162–177.

https://doi.org/10.1080/00461520.2022.2089147

Matsumoto, D., & Hwang, H. S. (2013). Cultural Influences on Nonverbal Behavior. In D. Matsumoto, M. Frank, & H. Hwang, *Nonverbal Communication: Science and Applications* (pp. 97–120). SAGE Publications, Inc. https://doi.org/10.4135/9781452244037.n5

Mayer, J. D. (2004). Christian Fundamentalists and Public Opinion Toward the Middle East: Israel's New Best Friends? *Social Science Quarterly*, *85*(3), 695–712.

Mayer, R. E. (2002). Multimedia learning. In *Psychology of Learning and Motivation* (Vol. 41, pp. 85–139). Academic Press. https://doi.org/10.1016/S0079-7421(02)80005-6

Mayer, R. E. (2014). Cognitive Theory of Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 43–71). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.005

Mayer, R. E., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, *38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

Mäyrä, F. (2008). *An introduction to game studies: Games in culture*. SAGE.

McCall, J. (2016). Teaching History With Digital Historical Games: An Introduction to the Field and Best Practices. *Simulation & Gaming*, *47*(4), 517–542. https://doi.org/10.1177/1046878116646693

McCall, J. (2023). *Gaming the Past: Using Video Games to Teach Secondary History* (2nd ed.). Routledge.

McConahay, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology*, *72*(2), 262–274. https://doi.org/10.1037/0022-3514.72.2.262

McConahay, J. B., Hardee, B. B., & Batts, V. (1981). Has Racism Declined in America?: It Depends on Who is Asking and What is Asked. *Journal of Conflict Resolution*, *25*(4), 563–579. https://doi.org/10.1177/002200278102500401

McCrudden, M. T., & Schraw, G. (2007). Relevance and Goal-Focusing in Text Processing. *Educational Psychology Review*, *19*(2), 113–139. https://doi.org/10.1007/s10648-006-9010-7

McKirnan, D. J., Smith, C. E., & Hamayan, E. V. (1983). A sociolinguistic approach to the belief-similarity model of racial attitudes. *Journal of Experimental Social Psychology*, *19*(5), 434–447. https://doi.org/10.1016/0022-1031(83)90021-5

Michael, D., & Chen, S. (2006). *Serious Games: Games That Educate, Train, and Inform*. Cengage Learning PTR.

Ministerium für Kultus, Jugend und Sport Baden-Württemberg. (2016). *Bildungsplan 2016: Geschichte (Gymnasium)*. https://www.bildungsplaene-bw.de

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, *91*(2), 358–368. https://doi.org/10.1037/0022-0663.91.2.358

Murayama, K., FitzGibbon, L., & Sakaki, M. (2019). Process Account of Curiosity and Interest: A Reward-Learning Perspective. *Educational Psychology Review*, *31*(4), 875–895. https://doi.org/10.1007/s10648-019-09499-9

Myers, D. G., & Twenge, J. M. (2022). *Social psychology* (Fourteenth edition). McGraw Hill Education.

Nacke, L. E., Grimshaw, M. N., & Lindley, C. A. (2010). More than a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game. *Interacting with Computers*, *22*(5), 336–343. https://doi.org/10.1016/j.intcom.2010.04.005

Naul, E., & Liu, M. (2020). Why Story Matters: A Review of Narrative in Serious Games. *Journal of Educational Computing Research*, *58*(3), 687–707. https://doi.org/10.1177/0735633119859904

Newzoo. (2025). *2025 Global Games Market Report* [Market Report]. Newzoo. https://newzoo.com/resources/trend-reports/newzoo-global-games-market-report-2025

Ninaus, M., Greipl, S., Kiili, K., Lindstedt, A., Huber, S., Klein, E., Karnath, H.-O., & Moeller,

K. (2019). Increased emotional engagement in game-based learning – A machine learning approach on facial emotion detection data. *Computers & Education*, *142*, 103641. https://doi.org/10.1016/j.compedu.2019.103641

Nosek, B. A. (n.d.). *Implicit Association Test (IAT): Race (Black/White) IAT*. Project Implicit. http://projectimplicit.net/nosek/iat/

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G., & Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*, 36–88. https://doi.org/10.1080/10463280701489053

Novo-Corti, I., Muñoz-Cantero, J.-M., & Calvo-Porral, C. (2011). Analysis of young college students' attitudes toward physical handicaps: A theory of planned behavior view [in Spanish]. *RELIEVE - Revista Electrónica de Investigación y Evaluación Educativa*, *17*(2). https://doi.org/10.7203/relieve.17.2.4010

OECD. (2015). Students, Computers and Learning: Making the Connection. *PISA*. https://doi.org/10.1787/9789264239555-en

OECD. (2021). 21st-Century Readers: Developing Literacy Skills in a Digital World. *PISA*. https://doi.org/10.1787/a83d84cb-en

Ohanian, R. (1990). Construction and Validation of a Scale to Measure Celebrity Endorsers' Perceived Expertise, Trustworthiness, and Attractiveness. *Journal of Advertising*, *19*(3), 39–52. https://doi.org/10.1080/00913367.1990.10673191

Paolini, S., Gibbs, M., Sales, B., Anderson, D., & McIntyre, K. (2024). Negativity bias in intergroup contact: Meta-analytical evidence that bad is stronger than good, especially when people have the opportunity and motivation to opt out of contact. *Psychological Bulletin*, *150*(8), 921–964. https://doi.org/10.1037/bul0000439

Papastergiou, M. (2009). Digital Game-Based Learning in high school Computer Science education: Impact on educational effectiveness and student motivation. *Computers & Education*, *52*(1), 1–12. https://doi.org/10.1016/j.compedu.2008.06.004

Park, J., Felix, K., & Lee, G. (2007). Implicit Attitudes Toward Arab-Muslims and the Moderating Effects of Social Information. *Basic and Applied Social Psychology*, *29*(1), 35–45. https://doi.org/10.1080/01973530701330942

Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour*, *5*(11), 1535–1547. https://doi.org/10.1038/s41562-021-01117-5

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*(3), 277–293. https://doi.org/10.1037/0022-3514.89.3.277

Pech, G., & Caspar, E. (2024). Can a Video Game with a Fictional Minority Group Decrease Intergroup Biases towards Non-Fictional Minorities? A Social Neuroscience Study. *International Journal of Human-Computer Interaction*, *40*(2), 482–496. APA PsycInfo (WOS:000866343700001). https://doi.org/10.1080/10447318.2022.2121052

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

Pekrun, R. (2024). Control-Value Theory: From Achievement Emotion to a General Theory of Human Emotions. *Educational Psychology Review*, *36*(3), 83. https://doi.org/10.1007/s10648-024-09909-7

Pekrun, R., & Linnenbrink-Garcia, L. (Eds.). (2014). *International Handbook of Emotions in Education* (0 ed.). Routledge. https://doi.org/10.4324/9780203148211

Peña, J., & Hernández Pérez, J. F. (2020). Game perspective-taking effects on willingness to help immigrants: A replication study with a Spanish sample. *New Media & Society*, *22*(6), 944–958. APA PsycInfo (2020-36417-002). https://doi.org/10.1177/1461444819874472

Peña, J., Pérez, J., Khan, S., & Gómez, A. (2018). Game Perspective-Taking Effects on Players' Behavioral Intention, Attitudes, Subjective Norms, and Self-Efficacy to Help

Immigrants: The Case of "Papers, Please." *Cyberpsychology, Behavior, and Social Networking*, *21*(11), 687–693. (WOS:000449924500004). https://doi.org/10.1089/cyber.2018.0030

Peng, W., Lee, M., & Heeter, C. (2010). The Effects of a Serious Game on Role-Taking and Willingness to Help. *Journal of Communication*, *60*(4), 723–742. https://doi.org/10.1111/j.1460-2466.2010.01511.x

Pettigrew, T. F., & Meertens, R. W. (1995). Subtle and blatant prejudice in western Europe. *European Journal of Social Psychology*, *25*(1), 57–75. https://doi.org/10.1002/ejsp.2420250106

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751–783. https://doi.org/10.1037/0022-3514.90.5.751

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist*, *50*(4), 258–283. https://doi.org/10.1080/00461520.2015.1122533

Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, *67*(4), 741–763. https://doi.org/10.1037/0022-3514.67.4.741

Prensky, M. (2001). *Digital game-based learning*. McGraw-Hill.

Ramasubramanian, S. (2010). Television viewing, racial attitudes, and policy preferences: Exploring the role of social identity and intergroup emotions in influencing support for affirmative action. *Communication Monographs*, *77*(1), 102–120. https://doi.org/10.1080/03637750903514300

Ramasubramanian, S. (2011). The Impact of Stereotypical Versus Counterstereotypical Media Exemplars on Racial Attitudes, Causal Attributions, and Support for Affirmative Action. *Communication Research*, *38*(4), 497–516. https://doi.org/10.1177/0093650210384854

Ravyse, W. S., Seugnet Blignaut, A., Leendertz, V., & Woolner, A. (2017). Success factors for serious games to enhance learning: A systematic review. *Virtual Reality*, *21*(1), 31–58. https://doi.org/10.1007/s10055-016-0298-4

Renninger, K. A., Bachrach, J. E., & Hidi, S. E. (2019). Triggering and maintaining interest in early phases of interest development. *Learning, Culture and Social Interaction*, *23*, 100260. https://doi.org/10.1016/j.lcsi.2018.11.007

Riopel, M., Nenciovici ,Lucian, Potvin ,Patrice, Chastenay ,Pierre, Charland ,Patrick, Sarrasin ,Jérémie Blanchette, & and Masson, S. (2019). Impact of serious games on science learning achievement compared with more conventional instruction: An overview and a meta-analysis. *Studies in Science Education*, *55*(2), 169–214. https://doi.org/10.1080/03057267.2019.1722420

Rosenthal, L., & Levy, S. R. (2010). The Colorblind, Multicultural, and Polycultural Ideological Approaches to Improving Intergroup Attitudes and Relations. *Social Issues and Policy Review*, *4*(1), 215–246. https://doi.org/10.1111/j.1751-2409.2010.01022.x

Rüsen, J. (2004). *History: Narration, interpretation, orientation*. Berghahn Books.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motivation and Emotion*, *30*(4), 344–360. https://doi.org/10.1007/s11031-006-9051-8

Sailer, M., & Homner, L. (2020). The Gamification of Learning: A Meta-analysis. *Educational Psychology Review*, *32*(1), 77–112. https://doi.org/10.1007/s10648-019-09498-w

Saleem, M., & Anderson, C. (2013). Arabs as Terrorists: Effects of Stereotypes Within Violent Contexts on Attitudes, Perceptions, and Affect. *Psychology of Violence, Technology and Violence*, *3*(1), 84–99. APA PsycInfo (WOS:000314745500007). https://doi.org/10.1037/a0030038

Saunders, C., Rutkowski, A. F., Genuchten van, M., Vogel, D., & Orrego, J. M. (2011). Virtual Space and Place: Theory And Test. *MIS Quarterly*, *35*(4), 1079–1098. https://doi.org/10.2307/41409974

Schachner, M. K., Schwarzenthal, M., Moffitt, U., Civitillo, S., & Juang, L. (2021). Capturing a nuanced picture of classroom cultural diversity climate: Multigroup and multilevel analyses among secondary school students in Germany. *Contemporary Educational Psychology*, *65*. https://doi.org/10.1016/j.cedpsych.2021.101971

Schumacher, A., Kammerer, Y., Scharinger, C., Gottschling, S., Hübner, N., Tibus, M., Kasneci, E., Appel, T., Gerjets, P., & Bardach, L. (2025). How do intellectually curious and interested people learn and attain knowledge? A focus on behavioral traces of information seeking. *European Journal of Personality*, 08902070241309124. https://doi.org/10.1177/08902070241309124

Schumacher, A., Kammerer, Y., Scharinger, C., Gottschling, S., Hübner, N., Tibus, M., Kasneci, E., Appel, T., Gerjets, P., & Bardach, L. (2026). How do intellectually curious and interested people learn and attain knowledge? A focus on behavioral traces of information seeking. *European Journal of Personality*, *40*(1), 87–110. https://doi.org/10.1177/08902070241309124

Schunk, D. H., Meece, J. L., & Pintrich, P. R. (2014). *Motivation in education: Theory, research, and applications* (4th edition, Pearson new international edition). Pearson.

Schwarzenthal, M., Schachner, M. K., Juang, L. P., & van de Vijver, F. J. R. (2020). Reaping the benefits of cultural diversity: Classroom cultural diversity climate and students' intercultural competence. *European Journal of Social Psychology*, *50*(2), 323–346. https://doi.org/10.1002/ejsp.2617

Seixas, P., & Morton, T. (2013). *The big six: Historical thinking concepts*. Nelson Education.

Serrano-Laguna, Á., Torrente, J., Moreno-Ger, P., & Fernández-Manjón, B. (2014). Application of Learning Analytics in educational videogames. *Entertainment Computing*, *5*(4), 313–322. https://doi.org/10.1016/j.entcom.2014.02.003

Shen, L. (2010). On a Scale of State Empathy During Message Processing. *Western Journal of*

*Communication*, *74*(5), 504–524. https://doi.org/10.1080/10570314.2010.512278

Shute, V., & Ventura, M. (2013). *Stealth Assessment: Measuring and Supporting Learning in Video Games*. The MIT Press. https://doi.org/10.7551/mitpress/9589.001.0001

Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*, *57*(10), 1380–1400. https://doi.org/10.1177/0002764213498851

Sitzmann, T. (2011). A Meta-Analytic Examination of the Instructional Effectiveness of Computer-Based Simulation Games. *Personnel Psychology*, *64*(2), 489–528. https://doi.org/10.1111/j.1744-6570.2011.01190.x

Siuko, J., Cloude, E., Paakkari, L., Dindar, M., & Kiili, K. (2025). Mobile Game vs. Digital Text: Gameplay Boosts Situational Interest, Satisfaction, and Emotional Engagement. *International Journal of Game-Based Learning (IJGBL)*, *15*(1), 1–21. https://doi.org/10.4018/IJGBL.398542

Spitzer, M. W. H. (2022). Just do it! Study time increases mathematical achievement scores for grade 4-10 students in a large longitudinal cross-country study. *European Journal of Psychology of Education*, *37*(1), 39–53. https://doi.org/10.1007/s10212-021-00546-0

Srivastava, S., Tamir, M., McGonigal, K. M., John, O. P., & Gross, J. J. (2009). The social costs of emotional suppression: A prospective study of the transition to college. *Journal of Personality and Social Psychology*, *96*(4), 883–897. https://doi.org/10.1037/a0014755

Stefaniak, A., Bilewicz, M., & Lewicka, M. (2017). The merits of teaching local history: Increased place attachment enhances civic engagement and social trust. *Journal of Environmental Psychology*, *51*, 217–225. https://doi.org/10.1016/j.jenvp.2017.03.014

Stephan, W. G., & Stephan, C. W. (1985). Intergroup Anxiety. *Journal of Social Issues*, *41*(3), 157–175. https://doi.org/10.1111/j.1540-4560.1985.tb01134.x

Strahan, R., & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlow-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, *28*(2), 191–193. https://doi.org/10.1002/1097-4679(197204)28:2%3C191::AID-JCLP2270280220%3E3.0.CO;2-G

Tamborini, R., Ron, Eastin, M., S., M., Skalski, Paul, Lachlan, K., Kenneth, Fediuk, A., T., Brady, & Robert. (2004). Violent Virtual Video Games and Hostile Thoughts. *Journal of Broadcasting & Electronic Media*, *48*, 335–357.

Tan, A., Fujioka, Y., & Tan, G. (2000). Television use, stereotypes of African Americans and opinions on affirmative action: An affective model of policy reasoning. *Communication Monographs*, *67*(4), 362–371. https://doi.org/10.1080/03637750009376517

Tan, A., Fujioka, Y., & Tan, G. (2009). Television use, stereotypes of African Americans and opinions on affirmative action: An affective model of policy reasoning. *Communication Monographs*, *67*(4), 362–371. https://doi.org/10.1080/03637750009376517

Teichman, Y., Bar-Tal, D., & Abdolrazeq, Y. (2007). Intergroup biases in conflict: Reexamination with Arab pre-adolescents and adolescents. *International Journal of Behavioral Development*, *31*(5), 423–432. https://doi.org/10.1177/0165025407081470

Thompson, M. M., Naccarato, M. E., Parker, K. C. H., & Moskowitz, G. B. (2001). The personal need for structure and personal fear of invalidity measures: Historical perspectives, current applications, and future directions. In *Cognitive social psychology: The Princeton Symposium on the Legacy and Future of Social Cognition* (pp. 19–39). Lawrence Erlbaum Associates Publishers.

Tobias, S., Fletcher, J. D., Bediou, B., Wind, A. P., & Chen, F. (2014). Multimedia Learning with Computer Games. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 762–784). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.037

United Nations. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development | Department of Economic and Social Affairs*. United Nations. https://sdgs.un.org/2030agenda

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

Van Looy, J., Courtois, C., De Vocht, M., & De Marez, L. (2012). Player Identification in

Online Games: Validation of a Scale for Measuring Identification in MMOGs. *Media Psychology*, *15*(2), 197–221. https://doi.org/10.1080/15213269.2012.674917

VandenBos, G. R. & American Psychological Association (Eds.). (2015). *APA dictionary of psychology* (Second Edition). American Psychological Association.

Vang, M., & Fox, J. (2014). Race in Virtual Environments: Competitive Versus Cooperative Games with Black or White Avatars. *Cyberpsychology, Behavior, and Social Networking*, *17*(4), 235–240. (WOS:000333766100006). https://doi.org/10.1089/cyber.2013.0289

Viccari, A., & Bardach, L. (2025). Designing Limes Text: Adapting an Educational Video Game into a Text-Based Format for Controlled Experimental Comparisons. *European Conference on Games Based Learning*, *19*(2), 855–863. https://doi.org/10.34190/ecgbl.19.2.3972

Viccari, A., Göllner, R., Hahn, J.-U., & Bardach, L. (2024). Limes (the Roman Frontier): Developing a Video Game for History Learning. *European Conference on Games Based Learning*, *18*(1), Article 1. https://doi.org/10.34190/ecgbl.18.1.2833

Von Keyserlingk, L., Krall, F., Krivošija, A., Yu, R., Lauermann, F., Arum, R., & Munteanu, A. (2025). *Perseverance Pays Off: Examining Behavioral Engagement in Learning Management Systems and Predicting Performance with Polygonal Curve Analysis*. PsyArXiv. https://doi.org/10.31234/osf.io/gcu96_v1

von Keyserlingk, L., Lauermann, F., Li, Q., Yu, R., Rubach, C., Arum, R., & Heckhausen, J. (2025). Students' study activities before and after exam deadlines as predictors of performance in STEM courses: A multi-source data analysis. *Learning and Individual Differences*, *117*, 102598. https://doi.org/10.1016/j.lindif.2024.102598

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Wineburg, S. (2001). *Historical Thinking & Other Unnatural Acts: Charting the Future of Teaching the Past*.

Winne, P. (2022). *Learning Analytics for Self-Regulated Learning* (pp. 78–85). https://doi.org/10.18608/hla22.008

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240. https://doi.org/10.1162/105474698565686

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, *105*(2), 249–265. https://doi.org/10.1037/a0031311

Wu, C.-H., Tzeng, Y.-L., & Huang, Y.-M. (2020). Measuring performance in leaning process of digital game-based learning and static E-learning. *Educational Technology Research and Development*, *68*(5), 2215–2237. https://doi.org/10.1007/s11423-020-09765-6

Yang, G., Gibson, B., Lueke, A., Huesmann, L., & Bushman, B. (2014). Effects of Avatar Race in Violent Video Games on Racial Attitudes and Aggression. *Social Psychological and Personality Science*, *5*(6), 698–704. APA PsycInfo (WOS:000338998600010). https://doi.org/10.1177/1948550614528008

Yu, Z., Yu, W. H., Fan, X., & Wang, X. (2014). An Exploration of Computer Game-Based Instruction in the “World History” Class in Secondary Education: A Comparative Study in China. *PLOS ONE*, *9*(5), e96865. https://doi.org/10.1371/journal.pone.0096865

Zavala-Rojas, D. (2014). Thermometer Scale (Feeling Thermometer). In *Encyclopedia of Quality of Life and Well-Being Research* (pp. 7199–7200). Springer, Cham. https://doi.org/10.1007/978-3-031-17299-1_1028