Multimodal Technologies and Interaction

MDPI

Article

Current State of Serious Games in Human Trafficking: Evaluation, Gaps, and Future Research Directions

Fakhroddin Noorbehbahani [1,*], Kingson Fung [2] and Kiemute Oyibo [2]

[1] Faculty of Computer Engineering, University of Isfahan, Isfahan 81746-73441, Iran
[2] P3 Department of Electrical Engineering and Computer Science, York University, Toronto, ON M3J 1, Canada; kingson2@yorku.ca (K.F.); kiemute.oyibo@yorku.ca (K.O.)
* Correspondence: noorbehbahani@eng.ui.ac.ir

Abstract: Addressing human trafficking is crucial due to its severe impact on human rights, dignity, and well-being. Serious games refer to digital games that are designed to entertain while also accomplishing at least one additional objective, such as learning or health promotion. Serious games play a significant role in raising awareness, training professionals, fostering empathy, and advocating for policy improvements related to human trafficking. In this study, we systematically examine and assess the current landscape of serious games addressing human trafficking to unveil the existing state, pinpoint gaps, and propose future research avenues. Our investigation encompassed academic publications, gray literature, and commercial games related to human trafficking. Furthermore, we conducted a thorough review of evaluation criteria and heuristics for the comprehensive assessment of serious games. Subsequently, incorporating these evaluation metrics and heuristics, the games were subjected to evaluation by both players and experts. Following a combined qualitative and quantitative analysis, the results were deliberated upon, and their implications were expounded. Five serious games related to human trafficking were identified and evaluated using the SGES and EGameFlow scales, along with both game-specific and serious game heuristics. Player and expert evaluations ranked "(Un)TRAFFICKED" and "Missing" as the best-performing games, while "SAFE Travel" received the lowest ratings. Players generally rated the games higher than experts, particularly in usability, feedback, and goal clarity, although the games scored poorly in audiovisual quality and relevance. Experts highlighted deficiencies in motivation, challenge, and learning outcomes. The lack of personalization and the absence of social gaming elements point to the need for more targeted human trafficking games adapted to different demographics, cultures, and player types.

Keywords: human trafficking; serious game; game evaluation; heuristic evaluation

Citation: Noorbehbahani, F.; Fung, K.; Oyibo, K. Current State of Serious Games in Human Trafficking: Evaluation, Gaps, and Future Research Directions. *Multimodal Technol. Interact.* **2024**, *8*, 95. https://doi.org/10.3390/mti8110095

Academic Editor: Johanna Pirker

Received: 9 September 2024
Revised: 5 October 2024
Accepted: 15 October 2024
Published: 24 October 2024

1. Introduction

The terms trafficking in persons and human trafficking (HT) are often used interchangeably as umbrella terms to describe criminal activities where traffickers abuse and profit from adults or children [1]. Trafficking involves taking control and ownership of individuals, treating them as property. Those who participate, directly or indirectly, aim to exploit others for their own gain, whether through forced labor, the sexual exploitation of adults or children, the removal of organs, and domestic servitude [2,3].

In the United States, two main forms of trafficking are recognized: forced labor and sex trafficking. Forced labor involves exploiting someone's services through force, fraud, or coercion. Domestic servitude is a type of forced labor where victims work in private residences, often in isolation. Forced child labor refers to schemes where traffickers compel children to work due to their vulnerability. Sex trafficking involves using force, fraud, or coercion to compel individuals into commercial sex acts, including exploiting children. Despite legal prohibitions and widespread condemnation, forms of slavery persist, such as the sale of children, forced child labor, and debt bondage [1].

Human trafficking is a pervasive and lucrative criminal activity worldwide. Based on some reports, human trafficking ranks as the third-largest criminal activity globally, following drug trafficking and counterfeiting [4]. Given the minimal expenses and substantial profits at stake, traffickers have a compelling motivation to persist in this abhorrent criminal activity [5].

Human trafficking generates an estimated annual global profit of $150 billion, victimizing around 25 million people worldwide [6]. According to the U.S. Department of Justice, a child is trafficked for sexual exploitation in the United States every two minutes [7].

Sexual exploitation is the most prevalent form of human trafficking, accounting for 79% of cases. The majority of victims of sexual exploitation are women and girls. Notably, in 30% of the countries that provided data on the gender of traffickers, women and girls constituted the largest group of traffickers. In certain regions, it is common for women to traffic other women [8]. The second most prevalent form of human trafficking is forced labor, representing 18% of cases, though this figure may be underestimated due to underreporting compared to trafficking for sexual exploitation. Globally, nearly 20% of all trafficking victims are children; however, in some areas of Africa and the Mekong region, children constitute the majority of victims, reaching up to 100% in parts of West Africa [8].

The global approach to combating human trafficking revolves around the "3P" paradigm—prosecution, protection, and prevention. This framework is endorsed by the United States, as evident in international agreements such as the Palermo Protocol and domestic legislation such as the Trafficking Victims Protection Act of 2000. The U.S. Department of State's Office to Monitor and Combat Trafficking in Persons (TIP Office) employs diplomatic and programmatic measures to promote the 3P paradigm worldwide. Additionally, a fourth "P" for partnership is recognized as a supplementary strategy to mobilize all segments of society in the fight against modern slavery [9].

Given its humanitarian implications, it is crucial to raise awareness and educate the public about human trafficking, not only to bridge the knowledge gap but, more importantly, to enhance the identification of victims and hold perpetrators accountable. Moreover, increased awareness can empower individuals to identify and report potential cases of human trafficking [5]. More importantly, it can lead to the early detection and education of potential victims, which can help them make informed decisions that can prevent them from being trafficked [10,11].

In particular, educational serious games hold prospects in tackling the scourge and prevalence of human trafficking, offering engaging tools that can raise widespread awareness and empower individuals. A serious game refers to a digital game designed to entertain while also accomplishing at least one additional objective, such as learning or health promotion. Although some equate serious games with educational games, digital games can serve "serious" purposes beyond learning. They can motivate individuals to exercise, be employed in medical treatment, or function as a marketing tool [12].

Serious games have become a promising educational method in diverse fields. For example, according to research conducted by Sharifzadeh et al. [13], serious games are increasingly used for health education. D'Errico et al. [14] investigate how playing a serious game impacts adolescents' perception of risks in home, school, and work environments. Results showed that playing the game increased engagement, the internal locus of control, risk perception, and protective behavioral intentions. Engagement and internal locus of control also acted as predictors for the other outcomes, highlighting the game's role in promoting safety and health awareness.

In the cybersecurity domain, most developed games focus on education, training, and raising awareness to enhance knowledge about cybersecurity [15]. For example, Phishy is an online serious game designed to train enterprise users in phishing awareness. It was shown that the Phishy game significantly enhances players' ability to identify phishing links while also providing an enjoyable gaming experience [16]. Gounaridou and colleagues present the development of a traffic safety educational game in which players follow road rules as pedestrians or drivers. The study demonstrates that well-

designed educational games can enhance engagement, improve traffic awareness, and foster social responsibility through experiential learning [17]. Additionally, serious games have been successfully applied in various educational fields, including science [18], circular economy [19], management [20], programming [21], cultural heritage [22], cognitive skill development [23], nursing education [24], etc.

In certain domains, more effort is needed to apply serious games to supplement traditional educational methods. For instance, while numerous apps assert to offer information on preventing child sexual abuse (CSA), the majority fall short on incorporating key features such as game-based learning or serious games for teaching children, involving parents in the education process, and providing age- and gender-specific education. The most effective methods for teaching children about sexual abuse prevention involve game-based approaches like gamification, game-based learning, and serious games [25].

Utilizing online games as a tool to raise awareness is an innovative approach. This method proves beneficial in educating individuals, particularly children and teenagers, about the intricacies of human trafficking in an interactive way. Through engaging in interactive games, users can familiarize themselves with various aspects and stages of human trafficking, ranging from recruitment, exploitation, and escape from trafficking rings to recovery, social reintegration, and the challenges faced in exercising the rights of trafficking victims. Employing video games becomes especially impactful when educating a younger audience about the realities of human trafficking [5].

In this research, to bridge the gap in the extant literature, we thoroughly investigated and evaluated existing serious games related to human trafficking to illuminate the current state of the art, identify gaps, and suggest future research directions. Specifically, we conducted an investigation into academic publications, gray literature, and commercial games related to human trafficking. Additionally, a comprehensive review of evaluation criteria and heuristics for assessing serious games was undertaken. After reviewing and incorporating evaluation metrics and heuristics, the games underwent evaluation by both players and experts. Following both qualitative and quantitative analyses, the results are discussed, and their implications are presented.

The research questions addressed in this paper are as follows:

- RQ1: What is the current state of publications on serious games relating to human trafficking?
- RQ2: How can existing human trafficking games be evaluated?
- RQ3: What are the outcomes and insights derived from the evaluation of serious games addressing human trafficking?
- RQ4: What are the gaps in the current serious game landscape related to human trafficking?
- RQ5: What future research directions should be explored to advance the field of serious games in the context of human trafficking?

The paper is organized as follows: Section 2 reviews and discusses academic publications related to human trafficking and serious games that address this issue. Additionally, this section provides a comprehensive exploration of serious game evaluation criteria and heuristics. Section 3 details the proposed game evaluation method, including the selection of human trafficking-related serious games, the determination of evaluation criteria, and the subsequent evaluation, examination, and analysis. Section 4 presents and discusses the results of both player-based and expert-based evaluations. Finally, Section 5 concludes the findings and outlines future research directions.

2. Related Work

2.1. Human Trafficking

Smith and colleagues [26] examine the scope of human trafficking, its negative impact on global society, and the relationship between human trafficking and corruption. The authors estimate that there are between 12 million and 30 million slaves worldwide, with

approximately 50% of trafficking victims being children and 70–80% being female. Their findings imply that human trafficking is a global problem that generates an estimated $32 billion in revenue annually, making it one of the most profitable crime industries in the world. The authors find a significant positive correlation between corruption and human trafficking, suggesting that countries with higher levels of corruption are more likely to experience higher levels of human trafficking. This study concludes that ending human trafficking requires changing people's attitudes and actions, as well as reducing corruption and increasing awareness of the issue.

The study by Khan et al. [27] reviews human trafficking prevalence in Asian countries, encompassing forced labor, forced marriage, and sex trafficking affecting men, women, and children. Analyzing 64 studies from 2015 to 2022, the authors identify key contributing factors such as poverty, unemployment, political instability, corruption, and natural disasters. The review indicates an estimated 40.3 million trafficking victims in Asia as of 2016, with 30% from South Asia, East Asia, and the Pacific. The authors emphasize the need for effective strategies and comprehensive legislation to address underlying causes. Recommendations include enhancing law enforcement, increasing public awareness, and improving socio-economic conditions to reduce trafficking risks.

Olisah et al. [28] present a thorough examination of human trafficking trends worldwide over a 20-year period. The study utilizes a robust dataset from the Counter Trafficking Data Collaborative (CTDC) and employs time-series analysis, predictive modeling, and data visualization techniques to identify patterns and trends in human trafficking. The research reveals a complex global landscape of human trafficking, with varying trends and patterns across different regions. The study identifies Africa, the Americas, Asia, and Europe as significant regions for human trafficking, with distinct patterns of exploitation and demographic vulnerabilities. The study emphasizes the need for targeted anti-trafficking efforts, cooperation among nations, and continuous research to combat human trafficking effectively.

The study conducted by Martin and her colleagues [29] shows that human trafficking is a significant problem in the United States, with demographic factors such as population, corruption, and religiosity playing a role in the prevalence of trafficking. The authors suggest that anti-trafficking efforts should focus on areas with high populations and high levels of corruption, and that education and awareness-raising efforts may be effective in reducing trafficking.

Albanese et al. [30] analyzed 27 studies involving interviews with over 3500 victims and offenders from 22 countries, highlighting the nuances of consent, coercion, and fraud in these relationships. They found that many adult victims consent to exploitative arrangements due to desperate situations, such as financial instability or family pressures. However, this consent is often tainted by coercion, manipulation, or deception. Moreover, coercion can be implicit, involving pressure, threats, or debt, rather than explicit physical force. Victims may be coerced by perpetrators, but also by circumstances, such as a lack of access to education, employment, or social services. This study shows that fraud is a common tactic used by traffickers to recruit and exploit victims. This can involve false promises of employment, education, or a better life, as well as the manipulation of victims' financial insecurity. The authors identify larger structural and social factors that contribute to human trafficking, including economic insecurity, housing insecurity, education gaps, and migration.

The paper by Saner et al. [31] discusses the challenges of measuring and monitoring human trafficking within the context of the 2030 Agenda and the Sustainable Development Goals (SDGs). It highlights the difficulties in collecting data on an invisible crime like human trafficking, which is often intertwined with issues like poverty, injustice, and weak institutions. The paper concludes that human trafficking, as an under-monitored issue, requires urgent attention and innovative solutions to improve data collection and policy responses. The authors recommend increasing awareness and citizen engagement in identifying trafficking.

Li et al. [32] propose a natural language processing method to identify potential human trafficking in massage business reviews. The authors created a keyword lexicon for human trafficking, building two classification models alongside BERT and Doc2Vec embeddings. Using a labeled dataset of Yelp reviews, they applied preprocessing techniques such as contractions, spelling corrections, and stop word removal. The models aim to automate the review screening process, reducing manual efforts by law enforcement. The study demonstrates the potential of natural language processing techniques in detecting human trafficking.

The current approach to addressing human trafficking, which focuses on due diligence and reporting, is insufficient, and a more holistic approach is needed. The limitations of the current due diligence approach include its focus on first-tier suppliers and its failure to address the root causes of human trafficking. A social connection and political responsibility model have been proposed, which emphasizes the need for businesses to take responsibility for their role in perpetuating human trafficking. Businesses have a responsibility to go beyond due diligence and to take affirmative action to prevent human trafficking in their supply chains [2].

Hodkinson et al. [33] argue that the UK government's efforts to combat modern slavery are flawed and counterproductive due to its hostile environment and policies toward migrants. The authors contend that the Modern Slavery Act 2015 focuses too narrowly on the immediate act of coercion between the victim and the perpetrator, ignoring the broader structural factors that contribute to migrant vulnerability and exploitation. They argue that the UK's hostile environment policies, which aim to deter irregular migration, actually create conditions that facilitate forced labor and exploitation. The UK government's hostile environment policies are incompatible with its efforts to combat modern slavery, and a more comprehensive approach is needed to address the root causes of migrant vulnerability and exploitation. A range of policy interventions have been proposed, including the provision of safe and legal migration routes, the restoration of rights and protections for asylum seekers, and the strengthening of labor market regulations to prevent exploitation [33].

The study by Chamber and colleagues [34] discusses the importance of trauma-informed care for survivors of human trafficking, who often experience complex post-traumatic stress disorder (PTSD) and trauma-coerced attachment (TCA). The authors, who are healthcare providers, share their experience and observations from working with survivors of human trafficking at a medical safe haven (MSH). The study concludes that trauma-informed care is essential for survivors of human trafficking, and that a comprehensive approach that addresses physical, psychological, and psychosocial healthcare needs is necessary.

2.2. Serious Games on Human Trafficking

In this study, a comprehensive selection of academic databases was employed as primary sources to identify publications on serious games related to human trafficking. These databases included Google Scholar, Web of Science, Scopus, Springer, Elsevier, Wiley, and PubMed. Additionally, we considered publications that cited the extracted records. The searches were conducted using targeted search terms pertinent to the title, keywords, and abstract sections.

The inclusion criteria for this review required that research be published in English and retrieved through the established search query. In cases where multiple papers reported the same study, only the most recent versions were considered, including theses, derived papers, and extended journal articles. Conversely, the exclusion criteria eliminated studies unrelated to the research questions, articles not written in English, and non-peer-reviewed sources such as opinion pieces and non-scholarly articles to uphold the research's reliability and credibility. The search query was as follows:

((Human Traffick* OR Traffick* in persons OR Modern Slavery OR Sexual exploit* OR Enslave* OR Debt bond* OR Forced labor OR Domestic servitude OR Organ traffick* OR Child exploit* OR Child soldier* OR Sex traffick* OR Forced prostitut* OR Forced marriage

OR Forced Begg* OR Forced Criminal) AND (Serious game* OR Video game* OR Digital game* OR VR OR Virtual Reality OR Augmented Reality OR Simulation OR Educational game* OR Game-based learning OR Mobile game* OR Interactive game*))

A number of researchers have carried out some important work on serious games focused on addressing human trafficking.

Toftedahl et al. [35] delve into the design and reception of a serious game called Missing, available on Google Play, with the goal of raising awareness about trafficking and its societal impact. The focus is on analyzing player metrics and Google Play app store data to understand player reception, emphasizing three key contributions: highlighting the tension between a designer's intention and game mechanics in conveying the message, addressing the complexity of finding relevant reviews for the serious theme, and examining the tension between star ratings and review content. A noteworthy finding is that even negative reviews can contribute positively to fulfilling the game's intended purpose. A review analysis on Google Play indicates overall appreciation for the storyline, but challenges arise in finding game mechanics that comprehensively align with all narrative aspects. Players, in particular, face difficulties related to progression and encounter bugs that impact the game mechanics [35].

O'Brien and Berents [36] explore three online games released in the past five years aimed at increasing awareness about human trafficking. The analysis highlights the prevalence of persistent tropes portraying ideal victims without agency, emphasizing individualized issues over structural causes. Despite this trend, the diverse approaches employed by the games showcase the potential for nuanced storytelling and complexity within the realm of digital games.

The Cybersecurity Institute has embraced the challenge of developing an immersive anti-trafficking training program that goes beyond mere awareness education [37]. It is designed to assess the specific skillsets of law enforcement and first responders. This comprehensive program aims to integrate all aspects of "serious gaming" within the framework of law enforcement and humanitarian communication. Given the dynamic nature of trafficking, the program, known as ATVRIT, will adapt and incorporate new insights into trafficking tactics and typologies as they emerge from law enforcement, academia, and victims' services organizations. Future iterations of ATVRIT will continually enhance the simulation environment to accurately mirror the evolving nature of trafficking situations. The programmers of ATVRIT recognize the increasing demand not only for effective and precise training but also for the inclusion of reflexive, harm-reducing techniques, addressing implicit biases and stereotypes in programming [37].

The first three studies reviewed serious games focused on human trafficking, while the subsequent two studies involved the development of a serious game related to this issue.

Koney and colleagues [38] focus on the application of art therapy to manage trauma in children rescued from trafficking at the Volta Lake at the Touch-a-Life-Care-Centre in Ghana. The objectives include exploring existing therapies at the center, examining current intervention methods, and testing the efficacy of a game-based intervention for trauma management in children. Using a case study approach with questionnaires, observations, and interviews, the study designed a game intervention using Scratch software. Results indicate that the game-based intervention in art therapy positively impacts traumatized children, enhancing their concentration and sustaining their interest in art classes. Children at the Touch-a-Life-Care-Centre welcomed the new intervention. The study recommends incorporating the Game Intervention in Art Therapy into the school curriculum and advocates for the recruitment of art therapy specialists in public health facilities to enhance effective interventions and improve the well-being of children in clinical art therapy sessions.

A game designed to simulate the challenges faced in real-life escape and rescue operations was introduced by Sanchawala et al. [39]. Drawing on established principles from educational literature, the authors aim to create a transformative experience that enables players to comprehend the obstacles that victims encounter and to gain insight into their mindset and thought processes. Their evaluation focuses on the game's effectiveness

in educating players about socio-economic situations, cultural predicaments, and latent conditions influencing human trafficking. To assess learning, they employ a two-phase survey process, consisting of a pre-test gauging players' knowledge about the current state of human trafficking and a post-test where players rate the game's experience, gameplay, and educational effectiveness regarding the trafficking scenario. Social activists engaged in rescuing trafficked individuals tested and validated the game, recognizing its potential impact on raising awareness.

In response to RQ1, we summarize the review of publications on serious games in the field of human trafficking in Table 1.

Table 1. Publications on serious games related to human trafficking.

Title	Year	Publication	# of Citations	Research Type	Description	Present Game Screens	Evaluation
[35]	2018	*International Journal of Serious Games*	7	Reviews existing HT games	Explores the design and player reception of Missing (a serious game aiming to raise awareness about trafficking) and its societal impact.	Yes	Yes (app data analysis—no serious game objectives evaluation)
[38]	2019	*British Journal of Education, Learning and Development Psychology*	0	Proposes a game	Introduces and tests for the potency of a game-based intervention as an additional intervention for managing trauma in victims.	Yes	Yes (tests degree of understanding the game through interviews and conducts emotion testing)
[36]	2019	*Anti-Trafficking Review*	9	Reviews existing HT games	Reviews three recent online games (BAN, ACT, (UN)TRAFFICKED) designed to raise awareness about human trafficking.	No	No
[39]	2020	IEEE Conference on Games (CoG)	1	Proposes a game	Presents a game immersing the player in realistic scenarios to educate about socio-economic situations, cultural challenges, and underlying factors in human trafficking crimes.	Yes	Yes (evaluates the learning process by having two phases of survey, pre-test and post-test; measures game's efficacy in educating players)
[37]	2021	*Anti-Trafficking Review*	0	Reviews ATVRIT program	Describes Anti-Trafficking Virtual Reality Immersion Training (ATVRIT), an anti-trafficking training program for law enforcement and first responders.	No	No

In response to RQ1, the current serious games addressing human trafficking that we found in academic literature do not mention any game development process/framework information to explain how the games are developed. User-centered design for the development of serious games focused on human trafficking has not been applied [40,41]. Moreover, the reviewed games have not been thoroughly evaluated. The game presented by Borrelli and Greer [37] has not been evaluated. Serious games need to be evaluated based on both their serious and game components [42]. Only both parts of the Unlocked game [39] have been evaluated; however, the authors have not applied any statistical tests, standard usability metrics, or serious game evaluation heuristics. Koney et al. [38] only evaluated the emotions of the players, and again, the results have not been statistically tested. The human trafficking information provided by the current educational games is very limited. For example, Unlocked [39] only provides limited information about factors possibly hindering a victim's escape and how serious the situation of human trafficking is in India. Missing [35], ACT, BAN, and (UN)TRAFFICKED [36] have not been presented in academic research papers by their developers. We evaluate these commercial/non-academic serious games about human trafficking in this paper. We identified only five publications related to human trafficking games, of which two studies involved the development of a game, and the total number of citations is 17. This suggests a scarcity of academic research in the field.

2.3. Serious Game Evaluation

In this section, we review studies related to the evaluation of serious games. These studies are categorized into three main types: player-based evaluations, expert-based evaluations, and studies focusing on various evaluation methodologies, frameworks, and models. Each category is further explored in the following subsections.

2.3.1. Player-Based Evaluation Studies

Calderón and Ruiz provide a comprehensive summary of the current state of assessing serious games, drawing from a systematic literature review. The review identifies key assessment methods, application domains, game categories, features considered for educational effectiveness, assessment procedures, and participant population sizes. The research highlights that questionnaires and interviews are the predominant techniques for evaluating serious games. The primary quality characteristics assessed include game design, user satisfaction, usability, usefulness, understandability, motivation, performance, playability, pedagogical aspects, and user experience, among others [43].

Fu et al. introduce a comprehensive scale for evaluating user enjoyment in e-learning games, encompassing eight dimensions: immersion, social interaction, challenge, goal clarity, feedback, concentration, control, and knowledge improvement [44]. To validate the scale, four learning games from the university's online course "Introduction to Software Application" were employed as instruments. Survey questionnaires were distributed to course participants, resulting in 166 valid samples. The outcomes indicated satisfactory validity and reliability for the proposed scale, named EGameFlow.

While many researchers acknowledge serious games as effective tools for teaching and learning, the literature lacks cohesion and/or consensus concerning the factors that influence users' experiences and perspectives. Fokides et al. introduce a tool designed to assess a game's effectiveness while concurrently comparing user viewpoints [45]. The report details the creation and validation of a scale initially comprising seventy-two items distributed across thirteen factors. A total of 542 university students engaged in two serious games, with the administered questionnaire capturing their responses. The exploratory and confirmatory factor analysis determined that twelve factors and fifty-three items should be retained in the final scale.

Moizer et al. [46] aim to articulate and evaluate an approach to assess user experience within the framework of a dedicated serious game designed to meet the training requirements of individuals in social enterprises. Their paper details the creation of a survey in-

strument, rooted in an extensive literature review, to capture the multi-dimensional aspects of user experience. The evaluation process is elucidated, and findings from surveys conducted among individuals in social enterprises are analyzed and discussed. The outcomes underscore the effectiveness of the proposed method for evaluating user experience [46].

2.3.2. Expert-Based (Heuristic) Evaluation Studies

Heuristic evaluation stands out as the preferred method for assessing usability in games, especially when conducted by experts. Moreover, heuristics serve as design guidelines that are valuable tools for both designers and usability professionals.

The heuristic evaluation for playability (HEP) is an exhaustive set of heuristics designed for assessing playability [47]. These heuristics draw inspiration from the literature on productivity and playtesting, specifically tailored for the evaluation of video games, computer games, and board games. To gauge their face validity and effectiveness in comparison to traditional user testing methodologies, these heuristics were applied to an evolving game design. The findings indicate that HEP successfully identified qualitative similarities and differences when compared to user testing. Moreover, HEP proved most effective for evaluating general issues during the early phases of development, particularly with prototypes or mock-ups. When combined with user studies, HEP introduces a novel method for the HCI game community, contributing to the creation of more usable and playable games.

To render HEP heuristics applicable across various game genres and delivery methods, another study concentrates on a refined set known as the heuristics of playability (PLAY) [48]. Designed for early implementation in game development and to assist developers in the interim between formal usability/playability research phases, these heuristics were derived from effective scores on metacritic.com, a popular game review website. Fifty-four gamers assessed high- and low-ranked games against 116 potential heuristics. The study explores the implications of these heuristics in enhancing game quality, emphasizing their utility in design evaluation and self-report survey formats.

GAP is another set of principles focused on first-time players, tutorial use, and initial game play [49]. Results showed that heuristics are more effective than "unassisted intuition" not only in identifying problems but also in inspiring recommendations for enhancements to the games' player experience [49].

A book chapter outlines an approach to evaluate user experience in video games using heuristics [50]. The authors provide a concise overview of video games, introduce the concept of user-centered design for games, and delve into the history of heuristics for video games and the broader role of user experience in gaming. They propose a refined framework comprising two sets of heuristics (gameplay/game story, virtual interface) aimed at identifying critical issues in games. To assess its effectiveness in measuring user experience factors, they compare expert evaluations of six current games with user experience-based ratings from various game reviews. The findings suggest a correlation between the satisfaction of their framework and the average rating of the game.

The dual purpose of serious games, involving the simultaneous attainment of intended educational effects (the serious aspect) and entertainment value (the game aspect), is insufficiently addressed in existing studies on serious game evaluation. Caserman et al. sought to outline essential quality criteria for serious games [42]. The primary objective of this research is to identify crucial factors in serious games and to align existing principles and requirements from game-related literature to enhance the effectiveness and appeal of serious games. In addition to a literature review, workshop results are also incorporated. The authors propose quality criteria for both the serious and game aspects, with particular attention to maintaining a balanced integration between them.

The primary objective of the research conducted by Jerzak and Rebelo [51] was to delineate the essence of serious games and the evaluation process. The authors leveraged existing heuristics for games, along with their inherent weaknesses and strengths. They

consolidated and presented the most crucial heuristic elements for games, forming three sets of heuristic evaluations to pinpoint areas of convergence.

In another research, authors have synthesized diverse heuristics into a succinct framework for gaming enjoyment, organized around the concept of flow [52]. Flow, a widely acknowledged enjoyment model, comprises eight elements that encapsulate various heuristics found in the literature. The proposed model, GameFlow, delineates eight elements: concentration, challenge, skills, control, clear goals, feedback, immersion, and social interaction. Each element is accompanied by a set of criteria for attaining enjoyment in games. To initiate the exploration and validation of the GameFlow model, expert reviews were conducted on two real-time strategy games—one highly rated and one poorly rated—using the GameFlow criteria. This process yielded a more profound comprehension of enjoyment in real-time strategy games, shedding light on the strengths and weaknesses of the GameFlow model as an evaluation tool.

Existing evaluation methods for games are inadequate in assessing serious games due to a lack of understanding and a failure to encompass the seriousness of the content [53]. As a response to this gap, Jerzak and Rebelo introduced the Heuristic Evaluation for Serious Games (HESG), which comprises three modules: Game Play, Entertainment and Usability, and Game Mechanics. Each module is designed for remote and autonomous use, allowing for flexibility to accommodate the specific requirements of designers and evaluators. The primary objective of the HESG is to establish a comprehensive and easily applicable tool for evaluating various types of serious games. The HESG has demonstrated its effectiveness as a versatile and accessible method, serving as a valuable tool for evaluating serious games designed for specific training purposes [53].

While numerous heuristic guidelines tailored to the specifics of games have been introduced, they often concentrate on particular subsets of games or platforms. In response to this limitation, Yanez-Gomez et al. [54] proposed a modular approach that involves classifying existing game heuristics using metadata. They also introduce a tool called MUSE (Meta-heUristics uSability Evaluation tool) for games. This tool enables the reconstruction of heuristic guidelines based on metadata selection, catering to the unique requirements of each evaluation case.

2.3.3. Evaluation Methodologies/Frameworks Studies

A literature gap in game-based learning (GBL) evaluation, arising from the inconsistent use of elements, is addressed in a study [55]. By establishing terminology and scope across four conceptual levels, the study systematically categorizes GBL evaluation elements based on scope, definition, and usage. Utilizing directed content analysis of GBL evaluation literature from a prior systematic review, the research dimensionalizes GBL and breaks it down into factors/sub-factors according to theoretical constructs. This results in a structured and clear pattern for educational game evaluation. The further codification of metrics and mapping of relationships among GBL dimensions contribute to a conceptual framework offering enhanced insights into the learning process with educational games, guiding focus areas and evaluation criteria.

Abdellatif and colleagues [56] divided the identified quality characteristics of serious games into primary and secondary categories based on their utilization in existing literature. A framework was then proposed to assess various dimensions of serious games by selecting and combining relevant quality characteristics. A programming serious game, Robocode, was chosen as a case study. In this study, the framework was applied as fifteen students at Queen's University Belfast played the game and evaluated different quality characteristics according to the proposed framework. The results indicated an overall positive evaluation of Robocode; however, the framework suggested certain changes to enhance the game's understandability, making it more accessible for users to play without the need for supervision or tutors.

Martinez et al. introduced the Gaming Educational Balanced (GEB) Model, addressing limitations in serious game evaluation [56]. Built upon the Mechanics, Dynamics, and

Aesthetics Framework and the Four Pillars of Educational Games Theory, the GEB Model offers a metric for assessing serious games and guiding their development. Tested with three indie serious games focused on mental health awareness, the evaluation highlighted that while gameplay was commendable, the integration of educational content was lacking. Statistical and machine learning validation of the GEB metric was performed, confirming its clarity and players' ability to evaluate it accurately.

Employing suitable mechanisms for gameplay experience (GX) evaluation and measurement facilitates the validation of positive gameplay experiences. Nacke and colleagues introduced an approach to formalize evaluative methods and outline a roadmap for implementing these mechanisms in the realm of serious games [57]. The authors advocate for a three-layer framework for GX, each layer accompanied by a range of measurement methodologies. They highlight the potential application of this framework in the domains of game-based learning and serious gaming, particularly in sports and health contexts.

Usability testing, though frequently overlooked in serious game development, holds significant importance, as issues in usability can significantly impact user experience and, consequently, the learning outcomes of serious games. Olsen et al. [58] offered serious game developers a streamlined approach to incorporate usability testing efficiently and effectively into their development process. The authors advocate a three-tiered assessment approach that includes not only traditional usability but also evaluations of playability and learning. The authors believe that learning is the primary objective of serious games, and enjoyment is often crucial in achieving usage goals; hence, their proposed approach provides step-by-step procedures and associated measures for assessing usability, playability, and learning outcomes concurrently during game development.

According to Caserman et al. [42], there are three primary types of procedures for serious game evaluation: simple, pre/post, and pre/post/post. In the simple procedure, authors conduct a session with the serious game, and after gameplay, evaluation mechanisms are provided to the players. The pre/post procedure involves two stages of evaluation, one before using the serious game and another after. This procedure is commonly employed by authors assessing the level of knowledge acquisition that players gained through the serious game. The pre/post/post procedure is similar to the pre/post procedure but includes an additional stage. This new stage occurs after a period of weeks or months from the end of the second stage, aiming to evaluate the retention of learned knowledge. The simple procedure stands out as the most prevalent evaluation method. A total of 55% of studies applied a population size of up to 40 people for serious game evaluation. Consequently, evaluations of serious games did not typically involve a large number of participants.

3. Method

To evaluate serious games about human trafficking, we followed the 4-step method depicted in Figure 1. In the first step, a list of serious games about human trafficking was prepared through an investigation of academic publications, a review of gray literature, and an exploration of related commercial games. In the next step, we selected serious game evaluation criteria based on metrics and heuristics that we thoroughly investigated. We determined player and expert evaluation criteria. Then, we conducted player-based and expert-based evaluations based on the selected criteria. Finally, we examined and analyzed the games quantitatively and qualitatively. In the following subsections, the evaluation steps are described in more detail.

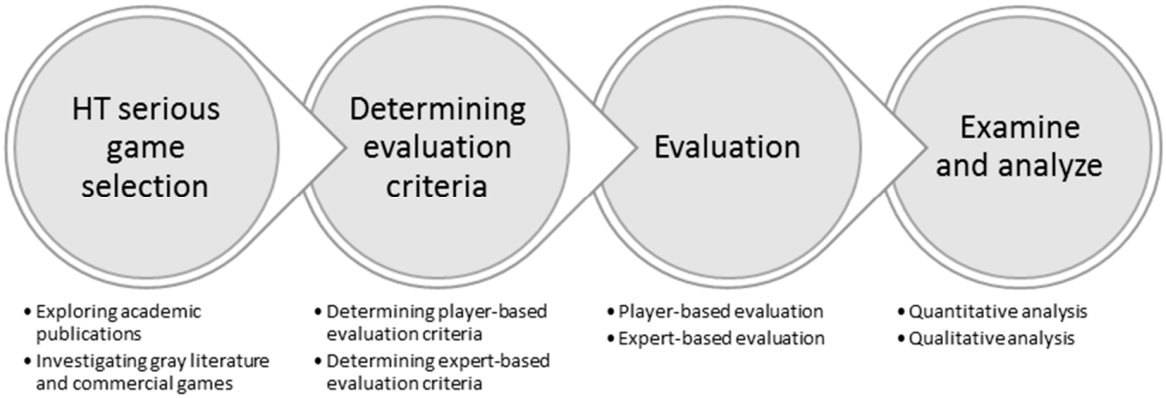
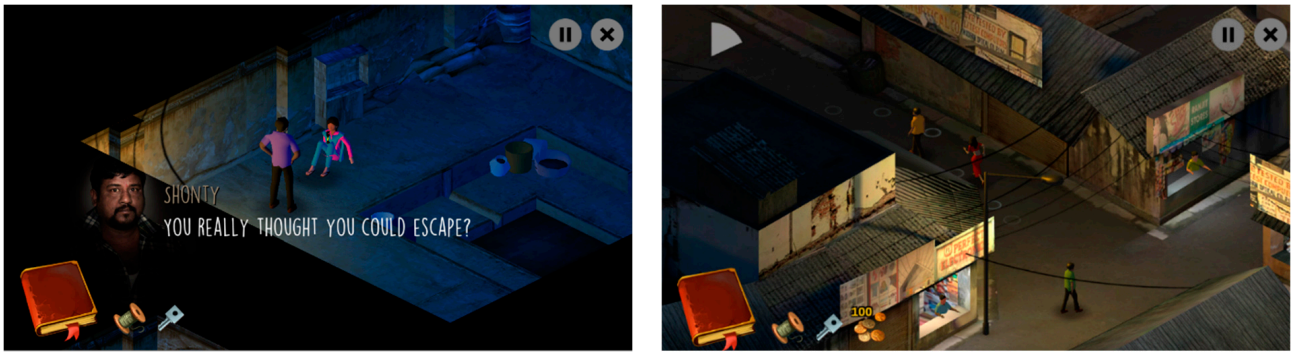


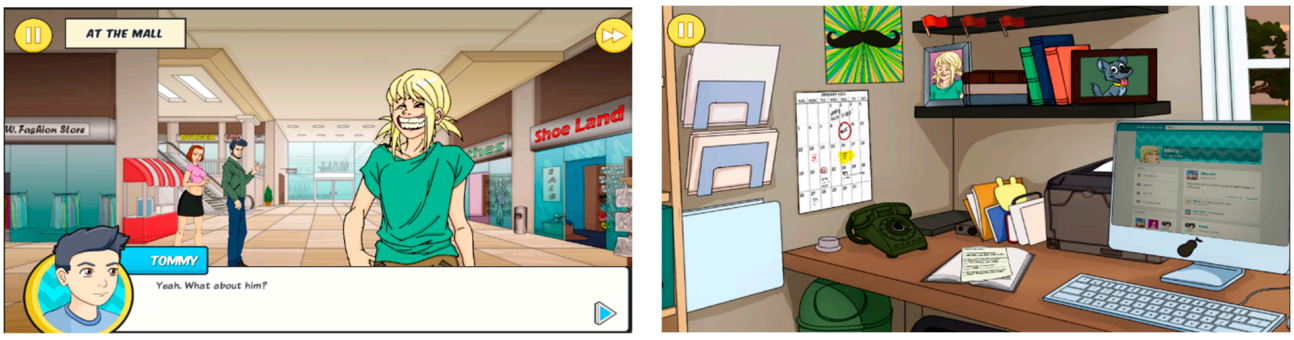
Figure 1. Game evaluation method, proposed by the authors.

3.1. HT Serious Games Selection

To find candidate HT serious games, we investigated academic publishing platforms, gray literature, and game publishing platforms such as Google Play, Appstore, Steam, and itch.io. It is notable that Unlocked [39], BeyondABC [59], and The Trap [60] are currently unavailable to the public. Dark Shadow [61] is still in development and has not been released to the public. After exploring games related to human trafficking that are currently available, we identified five prominent titles, as listed in Table 2. These games are accessible on platforms such as Google Play, the App Store, Steam, and other relevant websites. Figure 2 presents screenshots of the selected games.

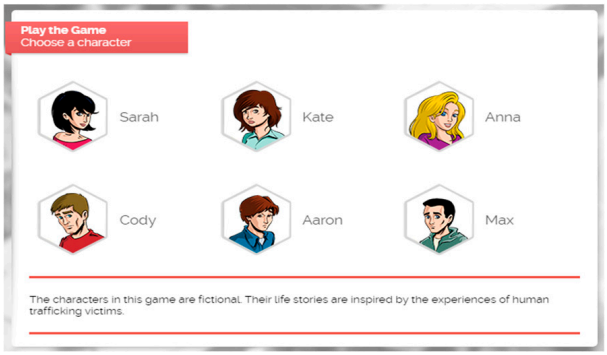


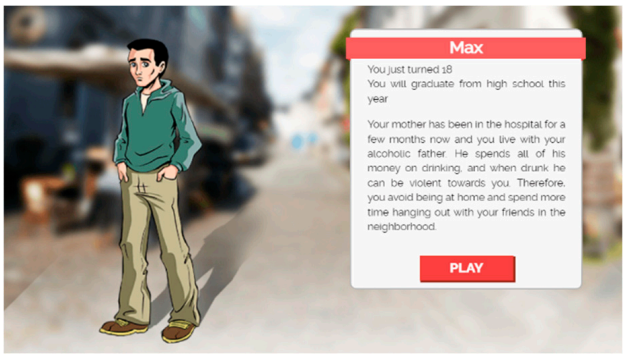
(**a**) Missing



(**b**) ACT

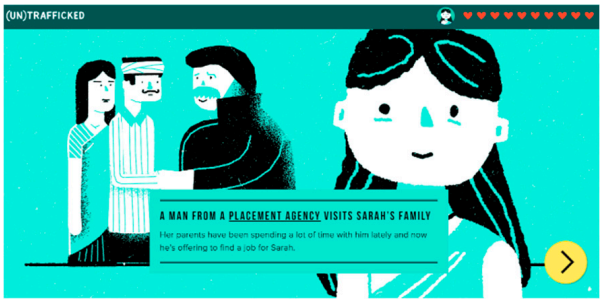
Figure 2. *Cont.*

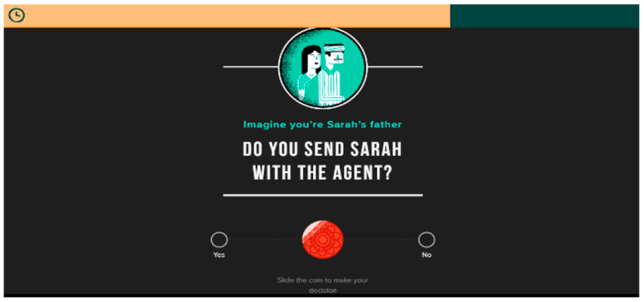




(**c**) BAN

(**d**) Safe Travel





(**e**) (UN)TRAFFICKED

Figure 2. Game screenshots.

Table 2. HT serious game specification.

Game	HT Type Covered	Game Objective	Game Mechanics	Audience	Genre	Platform	# of Players	Duration	# of Levels
Missing	Sexual trafficking	Raising awareness about sexual trafficking outcomes	Time limit, buying, levels	17+	AdventureRPG	Web, mobile, PC	1	≥3 h	4
ACT	Human trafficking	Educate about signs of human trafficking	Unlock, score, levels	13+	Visual novel	Web, mobile	1	≤1 h	3
BAN Human Trafficking!	Sexual exploitation, labor exploitation, forced marriage, forced criminality, domestic servitude, forced begging	Educate the users about the issue of human trafficking and the prevention of this type of crime	Narrative	Everyone	Visual novel	Web, mobile	1	≤20 min	-
SAFE Travel & Work Abroad	Sexual exploitation, labor exploitation, domestic servitude	Educate the users about the issue of human trafficking and the prevention of this type of crime	Narrative	13+	Visual novel	Web, mobile	1	≤10 min	-
(UN)TRAFFICKED	Sexual exploitation, labor exploitation	Raising awareness about human trafficking	Lives, narratives, time limit	Everyone	Visual novel	Web	1	≤10 min	-

3.2. Determining Evaluation Criteria

3.2.1. Player-Based Evaluation Criteria

In response to RQ2, to evaluate the games by players, after a thorough investigation of current metrics and scales, we integrated the EGameFlow [44] (56 items) and SGES [45] (53 items) scales. To accomplish this, we conducted a brainstorming session and compared the factors of the two scales. Upon reviewing the items of each factor, we found that some factors are common, with only variations in their names. For example, "immersion and presence" and "perceived learning effectiveness and knowledge improvement" are essentially the same.

As each player needs to evaluate 5 games, we aimed to streamline the scale by setting one item for each factor, combining related items based on the suggestions of three experts. Additionally, we excluded "perceived adequacy of the learning material" since the games under evaluation do not contain pedagogical materials and exercises. The "social interaction" factor was removed because there is no interaction between players during the evaluation of the games. Table 3 presents the 13 selected factors for evaluating serious games.

Table 3. Selected serious game evaluation factors.

Factor	Study		Selected
	EGameFlow [44]	**SGES [45]**	
Ease	Perceived ease of use	-	✓
Clarity	Perceived goal clarity	Goal clarity	✓
Autonomy	-	Autonomy	✓
Feedback	Perceived feedback	Feedback	✓
Audiovisual appeal	Perceived audiovisual adequacy	-	✓
Captivation	Perceived narratives	-	✓
Engagement	Motivation	Concentration	✓
Realism	Perceived realism	-	✓
Enjoyment	Enjoyment	-	✓
Challenge	-	Challenge	✓
Immersion	Presence	Immersion	✓
Learning	Perceived learning effectiveness	Knowledge improvement	✓
Relevance	Perceived relevance	-	✓
Social interaction	-	Social interaction	×
Perceived adequacy of the learning material	Perceived adequacy of the learning material	-	×

The final player-based evaluation 5-point Likert scale is displayed in Table 4, where PF_i stands for the ith player factor. Players also provided comments about the games in a free-text format.

Table 4. Player-based evaluation factors.

Factor	Item
PF1: Enjoyment	I think the game was fun and I enjoyed using the game.
PF2: Motivation	This game held my attention and motivated me to learn more about the learning subject.
PF3: Immersion	I forgot about time passing while using the game and felt detached from the outside world while using the game.
PF4: Ease of use/usability	I think it was easy to learn how to use the game, and it was easy for me to understand how to control the game.
PF5: Feedback	I received immediate feedback on my actions, and I was notified of new tasks immediately.
PF6: Narration/storyline	I was captivated by the game's story from the beginning and enjoyed the story provided by the game.
PF7: Perceived goal clarity	The game's goals were presented clearly, and the intermediate goals were presented at the beginning of each scene.
PF8: Audiovisual adequacy	I felt that the game's audio (e.g., sound effects, music) enhanced my (gaming) experience, and the game was visually appealing.
PF9: Realism	When interacting with the virtual objects, these interactions seemed like real ones.
PF10: Personal interest	The content of this game was relevant to my interests.
PF11: Usefulness	I felt that the game increased my knowledge and was a much easier way to learn compared to the usual teaching.
PF12: Autonomy	I felt a sense of control over the game.
PF13: Challenge	The difficulty of challenges increased as my skills improved, and the game provided new challenges with an appropriate pacing.

3.2.2. Expert-Based Evaluation Criteria

In response to RQ2, after thoroughly investigating serious game evaluation heuristics, we applied both serious and game-specific heuristics based on the research conducted by Polona et al. [42]. For the game-related aspect, we utilized Video Game Heuristics by Hochleitner et al. [50] because these heuristics are more comprehensive. Regarding the serious aspect, we applied all heuristics from Polona et al. [42], with the exception of the "quality" heuristic. This omission stems from our decision not to evaluate the games solely based on awards, ratings, and proof of effectiveness and sustainable effects. Instead, we opted to employ experts for the evaluation process. We included "regarding achieving serious goal" expression in both *progress feedback* and *reward* factors for added clarity (See Table 5).

Table 5. Expert-based evaluation factors (game part).

Category	Construct	Questions
Game play/ game story	EF1: Goal	• The game goals are clear. The game provides clear goals and presents overarching goals early as well as short-term goals throughout game play. • The skills needed to attain goals are taught early enough to play or use later or right before the new skill is needed.

Table 5. *Cont.*

Category	Construct	Questions
Game play/ game story	EF2: Motivation	• The player is receiving meaningful rewards. The acquisition of skills (personal and in-game skills) can also be a reward. • The game does not stagnate, and the player feels the progress. • The game play does not require the player to perform boring tasks. • Challenges are positive game experiences and encourage the user to continue playing. • The first-time experience is encouraging.
	EF3: Challenge	• The game is paced to apply pressure to but does not frustrate the player. • Challenge, strategy, and pace are in balance. • The challenge of the game is adapted to the acquired skills. The difficulty level varies so the player experiences greater challenges as they develop mastery. • The game is easy to learn but hard to master.
	EF4: Learning	• The player is given space to make mistakes, but the failure conditions must be understandable. • The learning curve is shortened. The user's expectations are met, and the player has enough information to get immediately started (or at least after reading the instructions once). • General help displaying the game's fundamentals exists, is a meaningful addition to the game, and provides useful assistance before and during the game. • Tutorials and adjustable levels are able to involve the player quickly (learning) and are provided upon request throughout the entire game.
	EF5: Control	• The player feels that they are in control. That includes the control over the character as well as the impact on the game world. It is clear what is happening in the game. • The player can impact the game world and make changes. • The player is able to skip non-playable and repeating content if not required by the game play. • The game mechanics feel natural and have correct weight and momentum. Furthermore, they are appropriate for the situation the player is facing. • The player is able to save the game in different states and is able to easily turn the game off and on.
	EF6: Consistency	• Changes the player makes to the game world are persistent and noticeable. • The game is consistent and responds to the user's action in a predictable manner. This includes consistency between the game elements and the overarching settings as well as the story.

Table 5. *Cont.*

Category	Construct	Questions
Game play/ game story	EF7: Story	• The meaningful game story supports the game play and is discovered as part of the game play. • The story suspends disbelief and is perceived as a single vision, i.e., the story is planned through to the end. • The game emotionally transports the player into a level of personal involvement (e.g., scare, threat, thrill, reward, punishment).
Virtual interface	EF8: Feedback	• The acoustic and visual effects arouse interest and provide meaningful feedback at the right time. • Feedback creates a challenging and exciting interaction and involves the player by creating emotions. • The feedback is given immediately to the player's action. • The player is able to identify game elements such as avatars, enemies, obstacles, power-ups, threats, or opportunities (orthogonal unit differentiation). • The player knows where they are on the mini-map if there is one and does not have to memorize the level design.
	EF9: Visual appearance	• In-game objects are standing out (contrast, texture, color, brightness) and cannot easily be misinterpreted. • The objects look like what they are intended to be (affordance).
	EF10: Interaction	• Input methods are easy to manage and have an appropriate level of sensitivity and responsiveness. • Alternative methods of interaction are available and intuitive. When existing interaction methods are employed, they are adhering to standards. • The first player action is obvious and results in immediate positive feedback.
	EF11: Customization	• The game allows for an appropriate level of customization concerning different aspects (e.g., audio and video settings, etc.). • The input methods allow for customization concerning the mappings. The customization is persistent.
	EF12: UI (user interface)	• The interface is consistent in control, color, typography, and dialogue design (e.g., large blocks of text are avoided, no abbreviations) and as non-intrusive as possible. • The menu is intuitive, and the meanings are obvious and perceived as a part of the game. • The visual representation (i.e., the view) allows the user to have a clear, unobstructed view of the area and of all visual information that is tied to the location. • Relevant information is displayed, and the critical information stands out. Irrelevant information is left out. The user is provided enough information to recognize their status and to make proper decisions. • If standard interface elements are used (buttons, scroll bars, pop-up menus), they are adhering to common game interface design guidelines.

For each game, the experts were requested to rate a total of 18 heuristics (displayed in Tables 5 and 6, where EF_i stands for the ith expert factor) using a 5-point Likert scale and considering the details of each heuristic. Additionally, the experts were encouraged to provide comments about the games in a free-text format.

Table 6. Expert-based evaluation factors (serious part).

Category	Construct	Questions
Serious goal	EF13: Serious goal focus	• Learning/training goal must remain in focus, for which a combination of physical and cognitive training can be beneficial. • Support players to achieve the serious goal. • Game elements should not interfere with the learning/training process.
	EF14: Clear goal	• Appropriate methods for the specific application area and target group. • Goals are clear and appropriate so that players can work toward the serious goal.
	EF15: Serious goal indispensability	• The serious part must be mandatory. • The serious goal must not be avoidable. • Training and learning tasks should not be a hurdle.
Methods	EF16: Content correctness	• Avoid errors and ensure that the content is technically correct. • Ensure correct technical language. • Remain neutral, especially on political and social issues.
	EF17: Progress feedback	• Players should receive feedback on their performance and progress regarding achieving the serious goal. • Visible and recognizable effects. • Provide simultaneous feedback (e.g., visual, audio, haptic, multimodal feedback).
	EF18: Reward	• Provide positive reinforcement and in-game awards regarding achieving the serious goal.

3.3. Evaluation

3.3.1. Player-Based Evaluation Procedure

Before the study began, the participants were given a survey that described the study's objectives and procedures. After agreeing to participate, they proceeded to the experiment. In total, five participants at a time played the games independently in the lab, each progressing at their own pace. Participants were first provided with tablets and asked to read the consent form, which introduced and provided an overview of the study. They were informed that the experiment aimed to evaluate the usability and effectiveness of five serious games focused on human trafficking. They also received the lead investigators' contact information in case they had any questions about the research. After reviewing this information, participants were asked to decide whether to give their consent to participate. All participants consented.

Those who consented played a portion of each of the five games and answered the same questionnaire for each game. Participants played only part of each game because some games required several hours to complete. The questionnaire included questions listed in Table 4, which participants answered using a 5-point Likert scale ranging from "strongly disagree—1" to "strongly agree—5." Additionally, participants recorded the start and end times for each game played. This information was used to calculate the time each participant spent playing each game in minutes.

The study, which evaluated five serious games on human trafficking, received approval from the second author's university ethics board. After approval, the study was announced on a course learning management system, allowing students to sign up for participation.

3.3.2. Expert-Based Evaluation Procedure

The experts were recruited through a targeted selection process based on their extensive experience in human–computer interaction (HCI), user experience (UX), and serious game expertise. We identified five candidates with strong backgrounds in these areas, focusing on those with published work or significant industry roles related to game design and evaluation. We reached out to these candidates via email and direct calls. Two of them accepted the invitation to collaborate, providing informed evaluations to ensure the credibility and reliability of our study. Initially, the evaluation criteria were explained to experts during an in-person meeting. Following this, they were asked to play the games and complete surveys provided via Google Forms. The experts were instructed to rate each game according to the evaluation heuristics outlined in Tables 5 and 6 using a 5-point Likert scale. Links to the questionnaires were sent to the experts, and they were asked to submit their responses.

3.4. Examine and Analyze

To undertake quantitative analysis, we utilized repeated measures ANOVA to assess whether there were significant differences in the mean ratings of the players across the five serious games. We additionally utilized ANOVA to examine whether there were significant differences in mean ratings corresponding to player-based evaluation factors across the various serious games. If the means exhibited significant differences, we conducted post-hoc analysis with a Bonferroni adjustment to determine which pair of games displayed significant distinctions. To present the true interval value of the mean rating produced by players for the games, we utilized a 0.95 confidence interval. In the case of expert evaluation, correlation analysis was employed to measure the agreement between experts' evaluations. We applied IBM SPSS Statistics Version 27 to conduct quantitative analysis.

In examining the data, we applied Braun and Clarke's 6-step approach to thematic analysis [62,63]. The process involves the following stages: familiarization, coding, theme development, reviewing themes, defining and naming themes, and writing up the themes. We initiated the analysis of open-ended questions through an inductive approach using semantic coding. Subsequently, we organized codes into categories and identified overarching themes across these categories. A member of the research team coded the data and engaged in discussions about themes and interpretations with another author during several meetings. We performed thematic analysis on the open-ended responses from both players and experts.

4. Results

4.1. Player-Based Evaluation Results

In the fall term of 2023, thirty-one students enrolled in a 300-level human–computer interaction (HCI) course at a Canadian university were recruited to participate in a study. This opportunity allowed them to experience a usability study from a participant's perspective, complementing their course requirement to conduct usability studies. As a token of appreciation, participants received a 2.5-mark bonus. The participants ranged in age from 15 to 30 and included 25 males and six females. In terms of ethnicity, 16 identified as Asian, 5 as Middle Eastern or North African, 3 as White, and 5 chose not to disclose their ethnicity. The research was approved by York University's Research Ethics Review Committee.

Figure 3 illustrates the mean ratings and confidence intervals of serious games addressing human trafficking (HT). The highest mean rating is associated with (UN)TRAFFICKED, while Safe Travel has the lowest mean rating. The results for ACT, Missing, and BAN demonstrate comparable mean ratings. Figure 4 displays mean player ratings corresponding to each factor across HT games. The mean values for immersion, audiovisual adequacy, realism, personal interests, and challenge are below 3.5, indicating that the games perform poorly in these factors and there is an opportunity to improve these features of serious games addressing human trafficking. The low value for personal interest suggests that players have little interest in human trafficking information, emphasizing the importance

of raising awareness and fostering interest in the subject. The mean values for enjoyment, motivation, narration/storyline, usefulness, and autonomy range between 3.56 and 3.96, indicating moderate ratings for the games in these factors, highlighting opportunities for enhancement. Finally, the mean values for ease of use, feedback, and goal clarity are above 4, signifying the strength of the games in these particular factors.

Figure 5 depicts the mean ratings associated with each factor for human trafficking (HT) games, and Table 6 presents the top- and bottom-performing games associated with each factor (numbers in parentheses represent the mean ratings). Concerning PF1 to PF7 and PF10 to PF12, (UN)TRAFFICKED outperforms the other games. The top-performing games for PF8, PF9, and PF13 are Missing, ACT, and Safe Travel, respectively. Regarding player-based evaluation, we can infer that (UN)TRAFFICKED is the best-performing game, while Safe Travel is the least favored. Table 7 highlights the highest- and lowest-performing games for each factor.

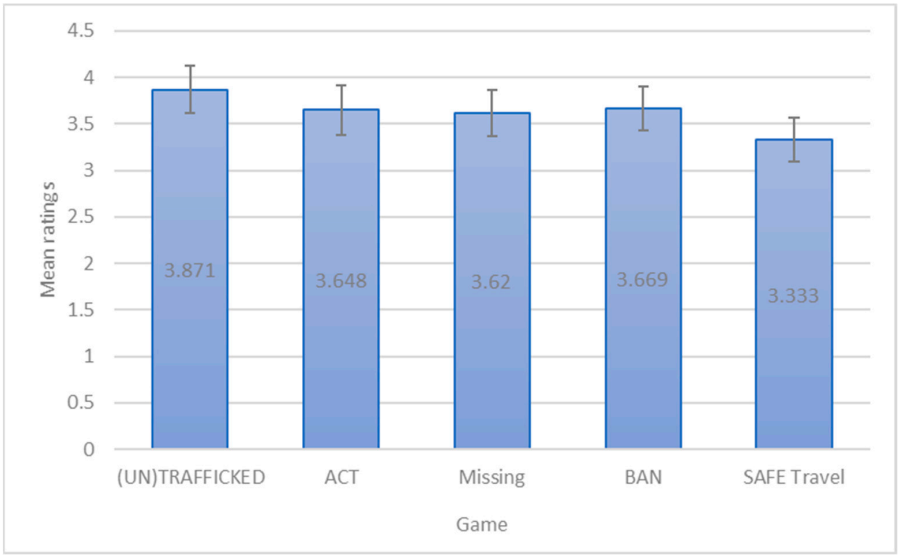


Figure 3. Mean player ratings of the HT games.

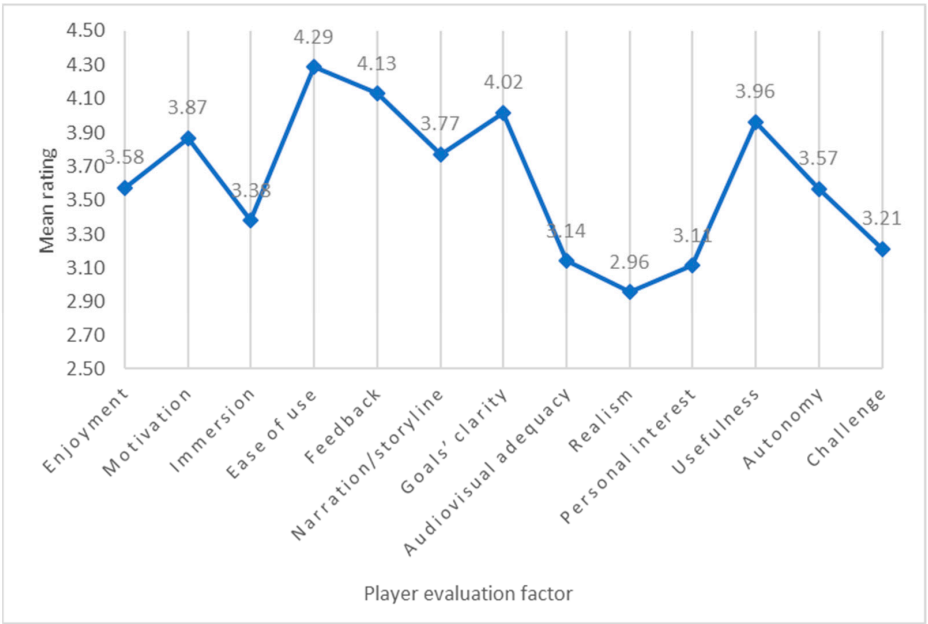


Figure 4. Mean player ratings corresponding to each factor across HT games.

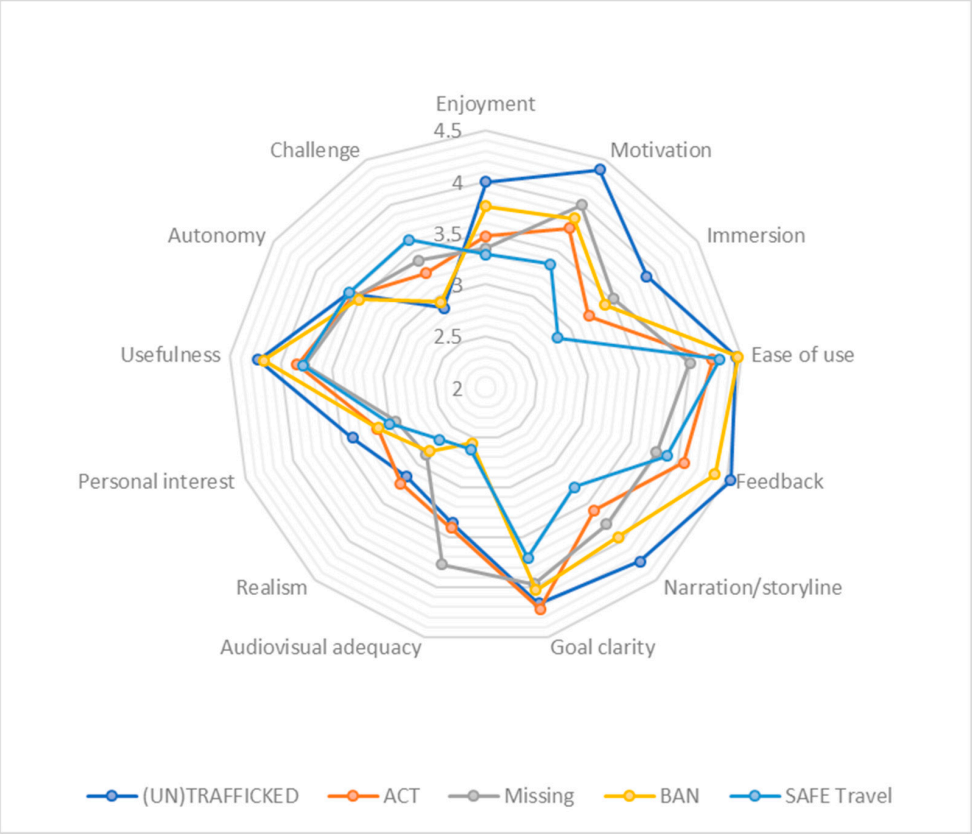


Figure 5. Mean player ratings associated with each factor.

Table 7. The top- and bottom-performing games associated with each factor.

Factor	Best	Worst
PF1: Enjoyment	(UN)TRAFFICKED (4.0)	SAFE Travel (3.29)
PF2: Motivation	(UN)TRAFFICKED (4.38)	SAFE Travel (3.35)
PF3: Immersion	(UN)TRAFFICKED (3.90)	SAFE Travel (3.85)
PF4: Ease of use/usability	(UN)TRAFFICKED (4.45)	Missing (4.0)
PF5: Feedback	(UN)TRAFFICKED (4.58)	Missing (3.77)
PF6: Narration/storyline	(UN)TRAFFICKED (4.25)	SAFE Travel (3.29)
PF7: Perceived goal clarity	(UN)TRAFFICKED (4.16)	SAFE Travel (3.70)
PF8: Audiovisual adequacy	Missing (3.77)	SAFE Travel (2.61)
PF9: Realism	ACT (3.25)	SAFE Travel (2.67)
PF10: Personal interest	(UN)TRAFFICKED (3.38)	(UN)TRAFFICKED (3.38)
PF11: Usefulness	(UN)TRAFFICKED (4.22)	Missing (3.77)
PF12: Autonomy	(UN)TRAFFICKED (3.77) Safe Travel (3.77)	BAN (3.5)
PF13: Challenge	SAFE Travel (3.61)	(UN)TRAFFICKED (2.87)

We conducted repeated measures ANOVA to compare the players' game ratings and PF1–PF13 values. Regarding games' ratings, PF1, and PF3–PF13, the assumption of sphericity was met, and therefore, no correction was applied to the degrees of freedom.

A repeated measures ANOVA determined that mean ratings differed statistically significantly between the games ($F(4, 120) = 2.751$, $p < 0.031$). Post-hoc analysis with a Bonferroni adjustment revealed that the mean game rating was statistically significantly

decreased from (UN)TRAFFICKED to Safe Travel 0.538 (95% CI, $p < 0.029$). Tables 8–10 show the details of ANOVA tests for comparing mean ratings of the games.

Table 8. Mauchly's test of sphericity [a].

Measure: Ratings							
Within-Subject Effect	**Mauchly's W**	**Approx. Chi-Square**	**df**	**Sig.**	**Epsilon [b]**		
					Greenhouse–Geisser	**Huynh–Feldt**	**Lower Bound**
Game	0.722	9.266	9	0.414	0.883	1.000	0.250

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix. [a] Design: intercept within-subject, design: game. [b] May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the tests of within-subject effects table.

Table 9. Tests of within-subject effects.

Measure: Ratings							
Source		**Type III Sum of Squares**	**df**	**Mean Square**	**F**	**Sig.**	**Partial Eta Squared**
Game	Sphericity assumed	4.602	4	1.150	2.751	0.031	0.084
	Greenhouse–Geisser	4.602	3.534	1.302	2.751	0.038	0.084
	Huynh–Feldt	4.602	4.000	1.150	2.751	0.031	0.084
	Lower bound	4.602	1.000	4.602	2.751	0.108	0.084
Error (Game)	Sphericity assumed	50.176	120	0.418			
	Greenhouse–Geisser	50.176	106.005	0.473			
	Huynh–Feldt	50.176	120.000	0.418			
	Lower bound	50.176	30.000	1.673			

Table 10. Pairwise comparisons.

Measure: Ratings						
(I) Game	**(J) Game**	**Mean Difference (I–J)**	**Std. Error**	**Sig. [b]**	**95% Confidence Interval for Difference [b]**	
					Lower Bound	**Upper Bound**
(UN)TRAFFICKED	ACT	0.223	0.195	1.000	−0.369	0.815
	Missing	0.251	0.161	1.000	−0.237	0.738
	BAN	0.202	0.168	1.000	−0.307	0.712
	SAFE Travel	0.538 *	0.166	0.029	0.036	1.041
ACT	(UN)TRAFFICKED	−0.223	0.195	1.000	−0.815	0.369
	Missing	0.027	0.159	1.000	−0.455	0.509
	BAN	−0.021	0.185	1.000	−0.580	0.538
	SAFE Travel	0.315	0.173	0.789	−0.210	0.840
Missing	(UN)TRAFFICKED	−0.251	0.161	1.000	−0.738	0.237
	ACT	−0.027	0.159	1.000	−0.509	0.455
	BAN	−0.048	0.163	1.000	−0.541	0.445
	SAFE Travel	0.288	0.121	0.241	−0.079	0.655

Table 10. *Cont.*

Measure: Ratings						
BAN	(UN)TRAFFICKED	−0.202	0.168	1.000	−0.712	0.307
	ACT	0.021	0.185	1.000	−0.538	0.580
	Missing	0.048	0.163	1.000	−0.445	0.541
	SAFE Travel	0.336	0.139	0.223	−0.086	0.758
SAFE Travel	(UN)TRAFFICKED	−0.538 *	0.166	0.029	−1.041	−0.036
	ACT	−0.315	0.173	0.789	−0.840	0.210
	Missing	−0.288	0.121	0.241	−0.655	0.079
	BAN	−0.336	0.139	0.223	−0.758	0.086

Based on estimated marginal means. * The mean difference is significant at the 0.05 level. [b] Adjustment for multiple comparisons: Bonferroni.

To avoid overwhelming the readers, we provide a summary of the results.

- A repeated measures ANOVA determined that mean *Enjoyment* differed statistically significantly between the games ($F(4, 116) = 3.565$, $p < 0.009$). Post-hoc analysis with a Bonferroni adjustment revealed that the mean of *Enjoyment* statistically significantly decreased from (UN)TRAFFICKED to Safe Travel by 0.80 (95% CI, $p < 0.021$).
- A repeated measures ANOVA determined that mean *Immersion* differed statistically significantly between the games ($F(4, 120) = 3.126$, $p < 0.017$). Post-hoc analysis with a Bonferroni adjustment revealed that the mean of *Immersion* statistically significantly decreased from (UN)TRAFFICKED to Safe Travel by 1.0 (95% CI, $p < 0.023$).
- A repeated measures ANOVA determined that mean *Feedback* differed statistically significantly between the games ($F(4, 120) = 3.328$, $p < 0.013$). Post-hoc analysis with a Bonferroni adjustment revealed that the mean of *Feedback* statistically significantly decreased from (UN)TRAFFICKED to Missing by 0.774 (95% CI, $p < 0.009$).
- A repeated measures ANOVA determined that mean *Narration/storyline* differed statistically significantly between the games ($F(4, 120) = 3.403$, $p < 0.011$). Post-hoc analysis with a Bonferroni adjustment revealed that the mean of *Narration/storyline* statistically significantly decreased from (UN)TRAFFICKED to Safe Travel by 0.968 (95% CI, $p < 0.010$).
- A repeated measures ANOVA determined that mean *Audiovisual adequacy* differed statistically significantly between the games ($F(4, 120) = 7.128$, $p < 0.001$). Post-hoc analysis with a Bonferroni adjustment revealed that the mean of *Audiovisual adequacy* statistically significantly decreased from Missing to BAN by 1.226 (95% CI, $p < 0.001$), and from Missing to Safe Travel by 1.194 (95% CI, $p < 0.001$).
- A repeated measures ANOVA determined that mean *Challenge* differed statistically significantly between the games ($F(4, 120) = 5.892$, $p < 0.001$). Post-hoc analysis with a Bonferroni adjustment revealed that the mean of *Challenge* statistically significantly decreased from (UN)TRAFFICKED to Missing by 0.903 (95% CI, $p < 0.005$) and from Missing to BAN by 0.806 (95% CI, $p < 0.024$).

Table 11 outlines the results of thematic analysis based on players' comments.

Table 11. Thematic analysis results of players' comments.

Game	Advantages	Disadvantages
(UN)TRAFFICKED	• Good story • Informative • Good graphics and sound	• Poor control • Display problem on android • Lack of onboarding
ACT	• Good story • Easy to play • Good graphics and sound	• Unclear process • Boring • Unclear interactive/non-interactive items • Need clear instruction
Missing	• Good story • Informative • Fun	• Poor movement • Too much dialogue • Poor control • Unnecessary challenges • Makes learning longer • Slow pace
Safe Travel	• Informative	• Boring • Very static • Poor control • Not a game • Super easy • Not engaging
BAN	• Informative • Good stories	• Poor control • Not engaging • Too wordy

4.2. Expert-Based Evaluation Results

Table 12 depicts the results of the correlation analysis conducted on expert ratings. A Pearson correlation of 0.849 indicates a strong correlation between experts' evaluations. The correlation is significant at the 0.01 level.

Table 12. Correlation analysis of experts' evaluation.

Correlations			
		E1	**E2**
E1	Pearson correlation	1	0.849 **
	Sig. (2-tailed)		0.000
	N	90	90
E2	Pearson correlation	0.849 **	1
	Sig. (2-tailed)	0.000	
	N	90	90

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

Figure 6 illustrates the expert ratings corresponding to serious games addressing human trafficking (HT). The highest mean rating is associated with Missing, while Safe Travel has the lowest mean rating. The mean scores of ACT, BAN, and Safe Travel are below 3.0, indicating poor performance of the games according to expert opinion. Figure 7 displays mean expert ratings corresponding to each factor across the HT games. The mean values for goal, motivation, challenge, learning, control, interaction, customization, progress feedback, and reward are below 3.0, indicating that the games perform poorly in

these factors. The mean value for serious goal indispensability is above 4, signifying the strength of the games in the indispensability of the serious goal. Regarding consistency, story, feedback, visual appearance, user interface, serious goal focus, clear goal, and content correctness, the mean values range from 3.1 to 3.6, indicating moderate ratings for the games in these areas and highlighting opportunities for improvement. Figure 8 depicts the experts' mean ratings associated with each factor for human trafficking (HT) games. Table 13 summarizes expert evaluation comments.

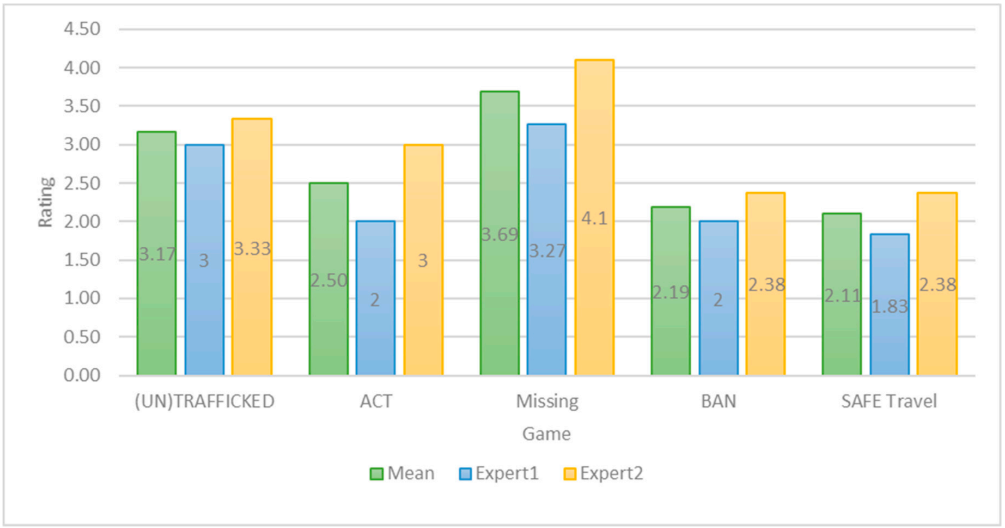


Figure 6. Expert ratings of the HT games.

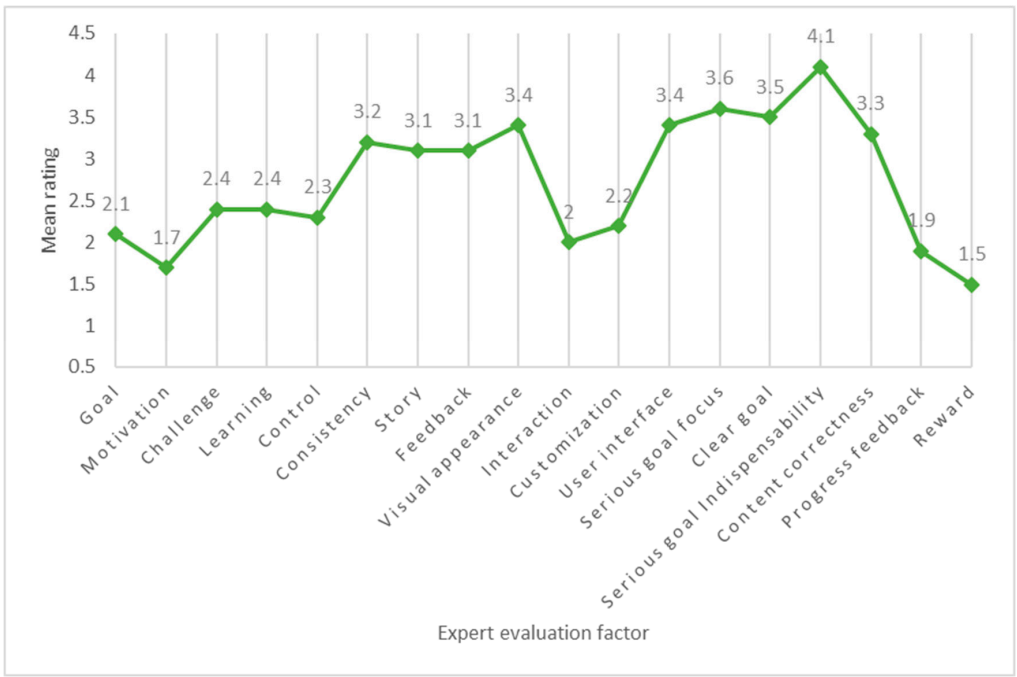


Figure 7. Mean expert ratings corresponding to each factor across HT games.

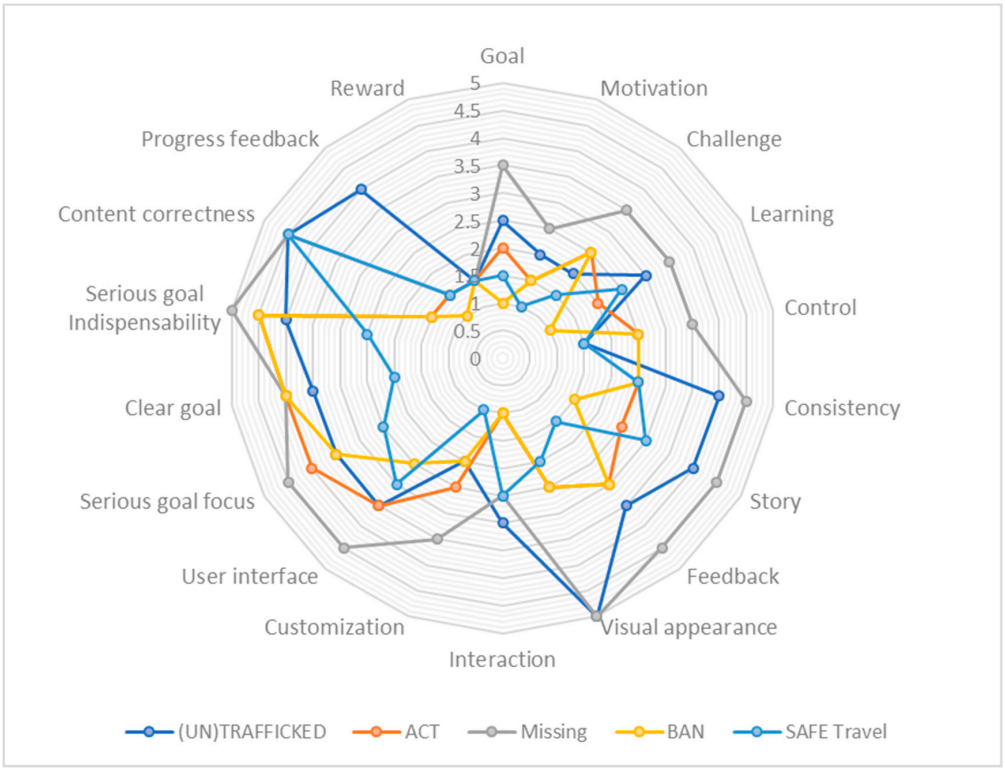


Figure 8. Mean expert ratings associated with each factor.

Table 13. Summary of expert evaluation comments.

Game	Advantages	Disadvantages
(UN)TRAFFICKED	• The sound and graphics are appealing. • The experience was good. The message was clear	• Game mechanics have not been applied to the game, and player could not control the game. For example, the hearts given to the player could not be maintained based on the player decisions. • Same as other games, the lack of using social mechanics to motivate the players and share HT information is evident. • The learning effectiveness of the player through the game have not been assessed, and the learning outcome is short term. • The HT information could be more comprehensive, including prevention, detection, and therapy tips.There was no menu or customization except the MC's name!
ACT	• I like the game and the idea of using red flags for determining human trafficking signs. • The quiz after each level reinforces the learning.	• The information provided in the game is limited to human trafficking signs. The preventive methods, therapy, rescue, and outcomes of trafficking have not been supported by the game. • Again, the lack of social game mechanics is evident. • The challenges are not adapted based on player skills, and in Act 2, the game stagnated. • There are no customization options.

Table 13. *Cont.*

Game	Advantages	Disadvantages
Missing	• Fascinating experience. • Visual and sound effect were helpful.	• The game includes too much dialogue, and in some levels, it is boring. • The sound and graphics are very good. There are some unnecessary game mechanics and interactive characters. For example, MEENA requests food and drink, but fulfilling her needs does not impact the game and it seems it is redundant. I could not find snacks and alcohol useful for coping with the challenges in the game and I think these elements are useless. • The game pace is slow, and it encompass boring and repetitive tasks. For example, every time the customer is served, the player should wait until the task is completed; this takes time, and the player does not have any choices and cannot skip. The process of making money is repetitive. • The game does not provide information for prevention, detection, or therapy respecting human trafficking, and it merely describes a specific scenario that highlights the outcomes of human trafficking. • The game does not provide any reward or feedback for learning goals. • As it is crucial to share information about human trafficking and informing other people, I suggest adding social game mechanics to the game.
Safe Travel	• The good thing about the app is that it contains several stories supporting different type of HT.	• I think this is not a game because it does not contain any game elements such as points, challenges, badges, leader boards, unlock, and so on. Therefore, this app could not support motivation, control, feedback, customization, etc. • The app is like a slideshow. • Total experience was not good! • The visuals were weak, plus there were no sound effects or music in background. • Player was not in control of the text and story pace. • No specific customization! • The rewards for the choices and consequences were unclear, and only in the end I found out about the situations and goals.
BAN	• Good game that improved my knowledge about human trafficking! • The stories were acceptable, although they were short.	• I think this is not a game because game mechanics are not used. • The player control is very limited, and only they see six stories of trafficking. • There is no audio, and the app is so boring and like a slideshow. • The effectiveness of the serious part is not measured by the game, and the short duration of the game seems to hinder long-term learning about human trafficking concepts. • It is crucial to apply social mechanics to the game for motivating the players and sharing HT information. • The HT information is limited to the outcome of HT, and it is necessary to incorporate prevention, detection, and therapy into the game.

4.3. Discussion

In response to RQ3, based on player and expert evaluations, "(UN)TRAFFICKED" and "Missing" were identified as the best games, respectively. Conversely, "SAFE Travel" was rated the worst by both players and experts. The mean ratings for all games were

3.61 for players and 2.73 for experts, indicating that players rated the games higher than experts did.

According to player evaluations, the games performed well in terms of usability, feedback, and the clarity and usefulness of perceived goals. However, they scored poorly on audiovisual adequacy, realism, relevance to personal interests, and challenge. The low rating for relevance to personal interests suggests that players have limited interest in information about human trafficking, underscoring the need to raise awareness and foster interest in the subject. The thematic analysis of player feedback highlighted issues such as low control, boredom, and low engagement as recurring themes.

From the expert perspective, the games received high ratings only for the indispensability of the serious goal. They were rated very low in areas such as goals, motivation, challenge, learning, control, interaction, customization, appropriate feedback on progress, and appropriate reward.

In response to RQ4, the lack of personalization and customization is evident in HT games, which could be tailored to individual player characteristics to improve effectiveness and user experience. Incorporating social game elements, such as inviting friends and multiplayer options, is vital for raising awareness about human trafficking and enhancing the player experience. Currently, these social game mechanics are absent from HT games. Additionally, it is crucial to develop HT serious games specifically designed for various demographics, including children, adolescents, males, females, parents, therapists, and law enforcement personnel.

Many of the games reviewed suffer from a lack of essential game mechanics. Players often have minimal control over the game, which undermines the interactive experience. The learning effectiveness of these games is questioned, with the impact being described as short term. The games fail to assess or reinforce the learning outcomes, which hinders the long-term retention of HT concepts. Furthermore, the educational content is often incomplete, focusing narrowly on human trafficking signs and outcomes without covering prevention, detection, therapy, or rescue. The games are criticized for their weak visuals and lack of sound effects or background music, which detracts from the immersive experience. Some games are so visually and audibly bland that they are compared to slideshows, failing to engage players on a sensory level.

There is a noticeable absence of reward systems or feedback mechanisms that could reinforce learning and motivate players. The unclear consequences of player choices and the lack of immediate feedback further reduce the effectiveness of these games in educating players about human trafficking. The overall user experience is described as poor, with several games failing to engage or interest the players. The combination of limited interactivity, slow pacing, and a lack of game elements like points, challenges, and badges contributes to this negative assessment.

In response to RQ5, to improve the effectiveness of serious games in combating human trafficking, several avenues for future work are proposed.

The future development of serious games should prioritize incorporating realistic scenarios and narratives that resonate with players, thereby increasing engagement and relevance. Personalization based on player preferences and characteristics, such as personality, culture, and player type, using models like the Hexad Player Type Model [64], can significantly enhance both the gaming experience and its educational impact. Adding social features, such as multiplayer modes and options to invite friends, can boost player interaction and expand the games' reach and effectiveness.

Additionally, developing games tailored to specific demographics—such as children, adolescents, men, women, parents, therapists, and law enforcement personnel—can improve the games' relevance and efficacy in educating diverse audiences. Efforts must also focus on raising awareness about human trafficking to enhance players' intrinsic interest and the perceived relevance of these games.

5. Conclusions and Future Work

This study assessed serious games designed to address the critical issue of human trafficking. We conducted a comprehensive investigation of both academic and gray literature to explore the landscape of HT serious games thoroughly. In addition, we examined player and expert evaluation criteria and proposed optimal evaluation metrics for these games. Our method, which combines player and expert evaluations, could be applied to assess other serious games.

In this study, we explored five key research questions related to serious games and human trafficking. First, we examined the current state of publications on serious games addressing human trafficking (RQ1). Next, we investigated how existing human trafficking games can be effectively evaluated (RQ2). We also analyzed the outcomes and insights derived from these evaluations (RQ3). Additionally, we identified gaps in the current serious game landscape related to human trafficking (RQ4). Finally, we proposed future research directions to advance the field of serious games in this context (RQ5).

Our study has highlighted a scarcity of academic publications on serious games related to human trafficking, with only five publications identified. This indicates a need for more research and development in this field. Existing human trafficking games lack thorough evaluation, particularly in applying user-centered design and comprehensive evaluation metrics. Serious games should be evaluated based on both their educational and entertainment components.

Quantitative and qualitative assessments were conducted using both player and expert participants, allowing us to identify the strengths and weaknesses of current game offerings. Notably, the game "(Un)TRAFFICKED" was preferred by players, while "Missing" was favored by experts, highlighting differences in evaluation criteria between these groups. Despite these differences, both groups agreed that "SAFE Travel" was the least effective game.

Players generally rated the games higher than experts, suggesting that while games are user-friendly and offer clear goals, they fall short in terms of realism, relevance, and challenge. The discrepancy highlights a critical gap between engaging gameplay and educational efficacy. Furthermore, the thematic analysis of players' comments revealed recurring issues such as a lack of control, low engagement, and uninteresting gameplay.

Experts rated the games highly only in terms of goal indispensability, with significant criticism directed at the games' ability to motivate, challenge, and educate. The lack of personalization and customization was a significant drawback, indicating that serious games need to be more adaptive to individual player needs and preferences.

The future development of serious games should focus on creating realistic scenarios and narratives to increase player engagement and relevance. Developing personalized serious games about human trafficking based on player type, culture, personality, and dominant persuasive strategies can enhance the gaming experience and educational effectiveness. Adding social elements such as multiplayer modes can improve interaction and broaden the game's impact. Moreover, designing games for specific groups—like children, adolescents, adults, parents, therapists, and law enforcement—can boost their effectiveness in educating varied audiences. Efforts should also aim to raise awareness about human trafficking to heighten players' interest and the perceived importance of these games.

This study has several limitations. The evaluation of games was based on partial gameplay rather than full engagement, which may have influenced the results. Engaging players in the full game could provide more comprehensive insights into user experience, motivation, and learning outcomes. Additionally, the sample size for player and expert evaluations may not fully represent the diverse demographics intended for these games. Future studies should consider longitudinal evaluations and larger, more diverse participant groups to obtain more generalizable findings.

Some of the survey questions that operationalized the user experience variables were double-barreled. This might have impacted participants' responses, as some participants might agree to one part of a question to a certain extent but not to the other

part. This must have made it difficult for the participants to decide and settle on a specific rating for the double-barreled questions. In future work, we plan to eliminate the double-barreled questions by streamlining and refining them to increase the reliability of participants' responses.

Author Contributions: Conceptualization, F.N. and K.O.; methodology, F.N. and K.O.; validation, F.N., K.F. and K.O.; investigation, K.F. and K.O.; data curation, F.N.; data analysis, F.N., K.F. and K.O.; writing—original draft preparation, F.N.; writing—review and editing, F.N. and K.O.; visualization, F.N.; supervision, K.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The research was approved by York University's Research Ethics Review Committee.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to ethical reasons.

Conflicts of Interest: The authors have no competing interests to declare.

References

1. US Department of State. Trafficking In Persons Report June 2023. Available online: https://www.state.gov/wp-content/uploads/2023/09/Trafficking-in-Persons-Report-2023_Introduction-V3e.pdf (accessed on 5 July 2024).
2. Ariail, D.L.; Smith, K.T.; Smith, L.M. Human Trafficking and Gender Inequality: How Businesses Can Lower Risks and Costs. *J. Risk Financ. Manag.* **2024**, *17*, 418. [CrossRef]
3. Cockbain, E.; Bowers, K. Human trafficking for sex, labour and domestic servitude: How do key trafficking types compare and what are their predictors? *Crime Law Soc. Chang* **2019**, *72*, 9–34. [CrossRef]
4. Assistant Attorney General Williams Delivers Remarks on the Department's Efforts to Fight Human Trafficking. Available online: https://www.justice.gov/olp/speech/assistant-attorney-general-williams-delivers-remarks-department-s-efforts-fight-human (accessed on 12 June 2023).
5. OSCE. Leveraging Innovation to Fight Trafficking in Human Beings: A Comprehensive Analysis of Technology Tools 2020. Organization for Security and Co-operation in Europe Vienna. Available online: https://www.osce.org/files/f/documents/9/6/455206_1.pdf (accessed on 19 June 2024).
6. *Countering Human Traffcking: Year in Review (October 2020 to September 2021)*; U.S. Department of Homeland Security, 2022. Available online: https://www.dhs.gov/sites/default/files/2024-03/24_0223_ccht_year-in-review-annual-report_508.pdf (accessed on 12 August 2024).
7. *Emerging from the Shadows (Human Trafficking Awareness Month Toolkit)*; Center for Prevention of Abuse, 2023. Available online: https://www.centerforpreventionofabuse.org/wp-content/uploads/2021/01/HTAM-2021-Toolkit.pdf (accessed on 12 August 2024).
8. United Nations UNODC. *Report on Human Trafficking Exposes Modern Form of Slavery*; United Nations, 2023. Available online: https://www.unodc.org/unodc/en/human-trafficking/global-report-on-trafficking-in-persons.html (accessed on 8 August 2024).
9. 3Ps: Prosecution, Protection, and Prevention. Available online: https://www.state.gov/3ps-prosecution-protection-and-prevention/ (accessed on 12 June 2023).
10. Rizo, C.F.; Chesworth, B.R.; Franchino-Olsen, H.; Klein, L.; Villodas, M.L.; Martin, S.L.; Macy, R.J. The State of Programs for Educating Youth about Sex Trafficking in the United States: A Nationwide Scoping Scan Survey. *J. Hum. Traffick.* **2023**, *9*, 513–531. [CrossRef]
11. Awerbuch, A.; Gunaratne, N.; Jain, J.; Caralis, P. Raising awareness of human trafficking in key professional fields via a multidisciplinary educational approach. *Int. J. Hum. Rights Healthc.* **2020**, *13*, 159–169. [CrossRef]
12. Ralf, D.; Göbel, S.; Effelsberg, W.; Wiemers, J. Foundations, Concepts and Practice. In *Serious Games*; Springer: Cham, Switzerland, 2016.
13. Sharifzadeh, N.; Kharrazi, H.; Nazari, E.; Tabesh, H.; Edalati Khodabandeh, M.; Heidari, S.; Tara, M. Health education serious games targeting health care providers, patients, and public health users: Scoping review. *JMIR Serious Games* **2020**, *8*, e13459. [CrossRef]
14. D'Errico, F.; Cicirelli, P.G.; Papapicco, C.; Scardigno, R. Scare-Away Risks: The Effects of a Serious Game on Adolescents' Awareness of Health and Security Risks in an Italian Sample. *Multimodal Technol. Interact.* **2022**, *6*, 93. [CrossRef]
15. Noorbehbahani, F.; Taghiyar, A. myREACH: A serious game for measuring security awareness about ransomware. *Multimed. Tools Appl.* **2024**, 1–35. [CrossRef]

16. CJ, G.; Pandit, S.; Vaddepalli, S.; Tupsamudre, H.; Banahatti, V.; Lodha, S. Phishy-a serious game to train enterprise users on phishing awareness. In Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts, Melbourne, Australia, 28–31 October 2018; pp. 169–181.
17. Gounaridou, A.; Siamtanidou, E.; Dimoulas, C. A Serious Game for Mediated Education on Traffic Behavior and Safety Awareness. *Educ. Sci.* **2021**, *11*, 127. [CrossRef]
18. Ullah, M.; Amin, S.U.; Munsif, M.; Yamin, M.M.; Safaev, U.; Khan, H.; Khan, S.; Ullah, H. Serious games in science education: A systematic literature. *Virtual Real. Intell. Hardw.* **2022**, *4*, 189–209. [CrossRef]
19. Whalen, K.A.; Berlin, C.; Ekberg, J.; Barletta, I.; Hammersberg, P. 'All they do is win': Lessons learned from use of a serious game for Circular Economy education. *Resour. Conserv. Recycl.* **2018**, *135*, 335–345. [CrossRef]
20. López, F.R.; Arias-Oliva, M.; Pelegrín-Borondo, J.; Marín-Vinuesa, L.M. Serious games in management education: An acceptance analysis. *Int. J. Manag. Educ.* **2021**, *19*, 100517. [CrossRef]
21. Schez-Sobrino, S.; Vallejo, D.; Glez-Morcillo, C.; Redondo, M.Á.; Castro-Schez, J.J. RoboTIC: A serious game based on augmented reality for learning programming. *Multimed. Tools Appl.* **2020**, *79*, 34079–34099. [CrossRef]
22. Theodoropoulos, A.; Antoniou, A. VR Games in Cultural Heritage: A Systematic Review of the Emerging Fields of Virtual Reality and Culture Games. *Appl. Sci.* **2022**, *12*, 8476. [CrossRef]
23. Reynaldo, C.; Christian, R.; Hosea, H.; Gunawan, A.A.S. Using Video Games to Improve Capabilities in Decision Making and Cognitive Skill: A Literature Review. *Procedia Comput. Sci.* **2021**, *179*, 211–221. [CrossRef]
24. Thangavelu, D.P.; Tan, A.J.Q.; Cant, R.; Chua, W.L.; Liaw, S.Y. Digital serious games in developing nursing clinical competence: A systematic review and meta-analysis. *Nurse Educ. Today* **2022**, *113*, 105357. [CrossRef]
25. Tasnuva Pritha, S.; Tasnim, R.; Ashad Kabir, M.; Amin, S.; Das, A. A Systematic Review of Mobile Apps for Child Sexual Abuse Education: Limitations and Design Guidelines. *arXiv* **2021**, arXiv:2107.01596.
26. Smith, K.T.; Martin, H.M.; Smith, L.M. Human trafficking: A global multi-billion dollar criminal industry. *Int. J. Public Law Policy* **2014**, *4*, 293–308. [CrossRef]
27. Khan, Z.; Kamaluddin, M.R.; Meyappan, S.; Manap, J.; Rajamanickam, R. Prevalence, causes and impacts of human trafficking in Asian countries: A scoping review. *F1000Research* **2023**, *11*, 1021. [CrossRef]
28. Olisah, S.; Odooh, C.; Efijemue, O.; Obunadike, E.; Onwuchekwa, J.; Owolabi, O.; Akintayo, S.; Obunadike, C. Unveiling Global Human Trafficking Trends: A Comprehensive Analysis. *J. Data Anal. Inf. Process.* **2024**, *12*, 49–75. [CrossRef]
29. Martin, H.M.; Smith, L.M. Historical overview and demographic analysis of human trafficking in the USA. *Int. J. Public Law Policy* **2015**, *5*, 219–248. [CrossRef]
30. Jay, S.; Albanese, R.B.; Gadd, D. Consent, Coercion, and Fraud in Human Trafficking Relationships. *J. Hum. Traffick.* **2022**, *8*, 13–32. [CrossRef]
31. Saner, R.; Yiu, L.; Rush, L. The measuring and monitoring of human trafficking. *Public Adm. Policy* **2018**, *21*, 94–106. [CrossRef]
32. Li, R.; Tobey, M.; Mayorga, M.E.; Caltagirone, S.; Özaltın, O.Y. Detecting Human Trafficking: Automated Classification of Online Customer Reviews of Massage Businesses. *Manuf. Serv. Oper. Manag.* **2023**, *25*, 1051–1065. [CrossRef]
33. Hodkinson, S.N.; Lewis, H.; Waite, L.; Dwyer, P. Fighting or fuelling forced labour? The Modern Slavery Act 2015, irregular migrants and the vulnerabilising role of the UK's hostile environment. *Crit. Soc. Policy* **2021**, *41*, 68–90. [CrossRef]
34. Chambers, R.; Gibson, M.; Chaffin, S.; Takagi, T.; Nguyen, N.; Mears-Clark, T. Trauma-coerced Attachment and Complex PTSD: Informed Care for Survivors of Human Trafficking. *J. Hum. Traffick.* **2024**, *10*, 41–50. [CrossRef]
35. Toftedahl, M.; Backlund, P.; Engström, H. Missing: Understanding the reception of a serious game by analyzing app store data. *Int. J. Serious Games* **2018**, *5*, 3–22. [CrossRef]
36. O'Brien, E.; Berents, H. Virtual Saviours: Digital games and anti-trafficking awareness-raising. *Anti-Traffick. Rev.* **2019**, *13*, 82–99. [CrossRef]
37. Borrelli, D.; Greer, B.T. The Next Step: The California Cybersecurity Institute's Anti-Trafficking Virtual Reality Immersion Training. *Anti-Traffick. Rev.* **2021**, *17*, 154–160. [CrossRef]
38. Koney, J.N.A.; Boateng, B.; Koomson, E. Managing Trauma in Trafficked Children Using a Game Intervention at the Touch a Life Care Center. *Br. J. Educ. Learn. Dev. Psychol.* **2019**, *2*, 43–59.
39. Sanchawala, A.J.; Arun, A.; Sajnani, R.; Vemuri, K. Unlocked: A Game On Human Trafficking. In Proceedings of the 2020 IEEE Conference on Games (CoG), Osaka, Japan, 24–27 August 2020; pp. 415–422.
40. Braad, E.; Žavcer, G.; Sandovar, A. Processes and models for serious game design and development. In *Entertainment Computing and Serious Games*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 92–118.
41. Ávila-Pesántez, D.; Rivera, L.A.; Alban, M.S. Approaches for serious game design: A systematic literature review. *ASEE Comput. Educ. J.* **2017**, *8*, 1–11.
42. Caserman, P.; Hoffmann, K.; Müller, P.; Schaub, M.; Straßburg, K.; Wiemeyer, J.; Bruder, R.; Göbel, S. Others Quality criteria for serious games: Serious part, game part, and balance. *JMIR Serious Games* **2020**, *8*, e19037. [CrossRef] [PubMed]
43. Calderón, A.; Ruiz, M. A systematic literature review on serious games evaluation: An application to software project management. *Comput. Educ.* **2015**, *87*, 396–422. [CrossRef]
44. Fu, F.-L.; Su, R.-C.; Yu, S.-C. EGameFlow: A scale to measure learners' enjoyment of e-learning games. *Comput. Educ.* **2009**, *52*, 101–112. [CrossRef]

45. Fokides, E.; Atsikpasi, P.; Kaimara, P.; Deliyannis, I. Let players evaluate serious games. Design and validation of the Serious Games Evaluation Scale. *ICGA J.* **2019**, *41*, 116–137. [CrossRef]
46. Moizer, J.; Lean, J.; Dell'Aquila, E.; Walsh, P.; Keary, A.A.; O'Byrne, D.; Di Ferdinando, A.; Miglino, O.; Friedrich, R.; Asperges, R.; et al. An approach to evaluating the user experience of serious games. *Comput. Educ.* **2019**, *136*, 141–151. [CrossRef]
47. Desurvire, H.; Caplan, M.; Toth, J.A. Using heuristics to evaluate the playability of games. In Proceedings of the CHI'04 Extended Abstracts on Human Factors in Computing Systems, Vienna, Austria, 24–29 April 2004; Association for Computing Machinery: New York, NY, USA, 2004; pp. 1509–1512.
48. Desurvire, H.; Wiberg, C. Game usability heuristics (PLAY) for evaluating and designing better games: The next iteration. In Proceedings of the Online Communities and Social Computing: Third International Conference, OCSC 2009, Held as Part of HCI International 2009, San Diego, CA, USA, 19–24 July 2009; pp. 557–566.
49. Desurvire, H.; Wixon, D. Game principles: Choice, change & creativity: Making better games. In *Proceedings of the CHI'13 Extended Abstracts on Human Factors in Computing Systems*; Association for Computing Machinery: New York, NY, USA, 2013; pp. 1065–1070.
50. Hochleitner, C.; Hochleitner, W.; Graf, C.; Tscheligi, M. A heuristic framework for evaluating user experience in games. *Game User Exp. Eval.* **2015**, 187–206.
51. Jerzak, N.; Rebelo, F. Serious games and heuristic evaluation--the cross-comparison of existing heuristic evaluation methods for games. In Proceedings of the Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience: Third International Conference, DUXU 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, 22–27 June 2014; pp. 453–464.
52. Sweetser, P.; Wyeth, P. GameFlow: A model for evaluating player enjoyment in games. *Comput. Entertain.* **2005**, *3*, 3. [CrossRef]
53. Jerzak, N.; Rebelo, F. Heuristic Evaluation for Serious Games (HESG): Using HESG for Designing and Evaluating Serious Games. *Adv. Ergon. Des. Usability Spec. Popul. Part I HLM* **2014**, *18*, 159–168. [CrossRef]
54. Yanez-Gomez, R.; Font, J.L.; Cascado-Caballero, D.; Sevillano, J.-L. Heuristic usability evaluation on games: A modular approach. *Multimed. Tools Appl.* **2019**, *78*, 4937–4964. [CrossRef]
55. Tahir, R.; Wang, A.I. Codifying game-based learning: The league framework for evaluation. In Proceedings of the 12th European Conference on Game Based Learning (ECGBL 2018), Sophia Antipolis, France, 4–5 October 2018; pp. 4–5.
56. Martinez, K.; Menéndez-Menéndez, M.I.; Bustillo, A. A New Measure for Serious Games Evaluation: Gaming Educational Balanced (GEB) Model. *Appl. Sci.* **2022**, *12*, 11757. [CrossRef]
57. Nacke, L.E.; Drachen, A.; Göbel, S. Methods for evaluating gameplay experience in a serious gaming context. *Int. J. Comput. Sci. Sport* **2010**, *9*, 1–12.
58. Olsen, T.; Procci, K.; Bowers, C. Serious games usability testing: How to ensure proper usability, playability, and effectiveness. In Proceedings of the Design, User Experience, and Usability. Theory, Methods, Tools and Practice: First International Conference, DUXU 2011, Held as Part of HCI International 2011, Orlando, FL, USA, 9–14 July 2011; pp. 625–634.
59. Parsons Students Combat Sex Trafficking through Game Design. Available online: https://blogs.newschool.edu/tnschangemakers/2017/12/parsons-students-combat-sex-trafficking-through-game-design/ (accessed on 10 March 2024).
60. The Trap: Human Trafficking Digital Education Tool. Available online: https://www.mcss.gov.on.ca/htapp/ (accessed on 10 March 2024).
61. SMU Creating Human Trafficking Data "Warehouse". Available online: https://www.smu.edu/News/Research/SMU-human-trafficking-project (accessed on 10 March 2024).
62. Braun, V.; Clarke, V. *Thematic Analysis in Encyclopedia of Critical Psychology*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 1947–1952.
63. Caulfield, J. How to do thematic analysis | Step-by-Step Guide & Examples. *Scribbr* **2023**. Available online: https://www.scribbr.com/methodology/thematic-analysis/ (accessed on 31 May 2024).
64. Krath, J.; Altmeyer, M.; Tondello, G.F.; Nacke, L.E. Hexad-12: Developing and Validating a Short Version of the Gamification User Types Hexad Scale. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, Hamburg Germany, 23–28 April 2023; Association for Computing Machinery: New York, NY, USA, 2023.