

Spatial and Temporal Dimensions: 2D vs. 3D in Serious Games for IT Security

Eren Dede¹, Tamara Wanner¹, and Marcus Gelderie^{1*}

¹Aalen University of Applied Sciences, Beethovenstr. 1, 73430 Aalen, Germany

Abstract

Serious games are an emerging tool for raising cybersecurity awareness and training employees, including in organisations operating critical infrastructures (CI). However, the specific role of spatial game design – such as 2D user interface (UI) environments versus 3D first-person worlds – remains largely unexplored. In this study, we investigate how the spatial dimension of a serious game influences learning outcomes, perceived immersion, and learner motivation in the CI context. We developed two conceptually and content-wise identical games: (1) a 2D UI-based card game and (2) a 3D first-person game set in a virtual hospital environment. In a between-subjects design ($n = 159$; 2D $n = 77$, 3D $n = 82$), participants completed a pre-test, played either the 2D or the 3D game, and responded to an immediate post-test and an EGameFlow-based questionnaire. A second knowledge post-test followed one week later. Our results show that both games improve cybersecurity knowledge, but we find no significant differences between the 2D and 3D versions in terms of knowledge gain, immersion, or motivation. Statistical analyses included paired t-tests for pre-post comparisons, Welch's two-sample t-tests for between-group comparisons, and Cohen's d for effect sizes.

Keywords: serious games; cybersecurity awareness; employee training; workplace learning; 2D vs. 3D game design

Received on 05 June 2026; accepted on 24 September 2026; published on 12 October 2026

Copyright © 2026 Eren Dede *et al.*, licensed to EAI. This is an open access article distributed under the terms of the (CCBY-NC-SA4.0), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi:10.4108/eetct.13359

1. Introduction

The number of cyberattacks against companies and critical infrastructures worldwide is increasing constantly [1]. This trend continues despite a growing number of learning opportunities that raise awareness [2]. Such attacks can have serious consequences, particularly in critical infrastructure (CI) sectors that provide essential services such as energy supply, healthcare, or telecommunications. Companies and institutions suffer considerable financial damage as a result of cyberattacks [3].

One way to reduce the risk of such attacks is to train employees in cybersecurity. Studies from Tolossa (2023) and Shillair (2022) show that targeted learning opportunities can help to increase security awareness and therefore reduce the attack surface [4, 5]. People are one of the biggest gateways for cyberattacks [6], which is why effective training methods are crucial.

One promising method for teaching cybersecurity concepts is the use of serious games [7–10]. Serious games are educational games that aim to impart knowledge in a fun and interactive way. Various studies suggest that serious games could be superior to traditional learning methods such as lectures or teaching texts [11, 12]. Such games could play a particularly valuable role in areas such as IT and critical infrastructure, as they can simulate real-life scenarios and realistic threats. The effectiveness of serious games in cybersecurity has been studied extensively (e.g. [13–15]). Previous studies have compared serious games to other forms of learning. But the effects of the nature of the game on its effectiveness in cybersecurity, for example its spatial characteristics (e.g. 2D or 3D environment), are surprisingly unexplored.

Problem Statement: The key question we address in this paper is how serious games need to be designed in order to achieve maximum learning impact. A characteristic feature of a game is its spatial dimension

*Corresponding author. Email: marcus.gelderie@hs-aalen.de

(e.g. first-person 3D, 2D isometric, or UI-only). It is natural to assume that the spatial dimension has an influence on the effectiveness of a game as teaching material and on the motivation and engagement of players. For example, consider a 3D game where players respond to a security breach by moving to a server room and shutting down the server. Doing so takes time and illustrates the impact of the response time on incident damage. Moreover, when players respond to an incident by interacting with their objects in a virtual environment, they might feel more motivated by the game. We seek to address the following overarching research question: “How does the spatial dimension influence the effectiveness and immersiveness of the game?”

Research Gap: While studies from Kaplan (2022) and Zaharias (2017) suggest that 2D games are more effective in purely imparting knowledge [16, 17], other studies from Karavidas (2024) and Makransky (2021) show that 3D games are more motivating and enhance the gaming experience due to their immersive environment [18, 19]. Studies from Aksoy (2019), Carbonell-Carrera (2021) and Dan and Rainer (2017) report a clear advantage for 3D environments, especially when spatial imagination is required [20–22], while other studies from Ak (2017), Kaimara (2019) and Lane (2010) have not found significant differences between the two formats [23–25].

Despite the increase in research on serious games, there has been no systematic investigation of whether 2D or 3D educational games are more effective at teaching cybersecurity concepts. This research gap is of particular relevance for companies that want to use such games to train their employees, as both the effectiveness and the cost-benefit factor play an important role. Furthermore, Pixune (2024) states that the development of 3D games is potentially more complex and costly than that of 2D games [26], making a targeted analysis of their learning effectiveness even more necessary.

Contributions: The main contributions of our paper are as follows:

- We propose a systematic way of measuring the influence of the spatial dimension on the effectiveness and immersive properties of a game as a learning tool in cybersecurity.
- Building on this, we provide a systematic study of the effects of the spatial dimension on the effectiveness of game-based learning in cybersecurity.
- Finally, we provide a statistical evaluation of the results of this study.

Methodology and Results: We investigate the effects of the spatial dimension of an educational game on

learning success and immersion. The focus is on the comparison of 2D, UI-only games and 3D learning games with regard to their effectiveness in teaching cybersecurity concepts in the critical infrastructure (CI) sector.

To this end, we developed two games with very similar game mechanics, but very different spatial characteristics. Both are turn-based games in which the player invests into cybersecurity defenses and reacts to possible incidents that may occur. The objective is to maintain a high availability, while avoiding bankruptcy. One game is a UI-only card game, where defenses and reactions are modeled as playing cards that cost money. The other is a first-person game in a continuous (yet bounded) 3D virtual hospital. Incidents are observed in a “war room”. Proactive defenses and reactions to events involve visiting various suppliers and government agencies. Both games have the same factual content and convey the same topics. Consequently, the learning objectives (stated in section 2) are identical for both games.

In order to analyze the effectiveness and immersive properties of both games, we conducted a field study. We grouped participants into two non-overlapping groups. Each group played one of the two games. Both groups were quizzed on the knowledge just before and one week after playing their game. Additionally, both answered a survey designed to measure the immersive properties of the two games (immediately after playing).

Our results show that both games are effective at teaching cybersecurity competencies (paired t-test comparison of pre- and post-play results). Surprisingly, our results show no statistically significant benefit of using a 3D continuous first-person game over a pure UI-based game (Welsh’s two-sample t-test comparison of both improvement and overall score of 2D vs. 3D group). This result applies to both the effectiveness and (even more surprisingly) the immersiveness and motivational properties of the game. As a result, the higher effort of developing such games may not pay off in terms of increased learning success.

Outline Our paper is structured as follows: We first define our methodology. We describe the two games that we developed and define the stated learning objectives. Next, we state our research questions and hypotheses. We also describe the sample population for our study in this section. In section 3, we describe and analyze our results and provide a discussion of those results in section 4.

Related Work

Serious games have been widely explored as an effective tool for teaching cybersecurity concepts. Several studies present from Mostafa (2019), Awojana

(2019) and Angafor (2020) custom-developed learning games and confirm their usefulness in promoting security awareness and skills [10, 27, 28]. Other research from Karagiannis (2021), Bacud (2021), Roepke (2019) and Jin (2018) analyses existing games and gamified platforms, showing positive effects on learners' knowledge gain [8, 29–31]. However, Ajsing (2011), Tobias (2011) and Mayer (2014) criticize that many serious games lack clearly defined learning objectives and fail to support long-term retention of knowledge or skills [11, 32, 33].

Several existing learning games focus specifically on cybersecurity training. For instance, Karagiannis (2021) [29] evaluates a game aimed at raising phishing awareness, while Bacud (2021) [30] introduces a simulation environment for network defense strategies. Similarly, Roepke (2019) [31] describes a role-playing game where players act as system administrators under attack, learning to prioritize risks and allocate limited resources. These works support the idea that interactive learning environments can enhance security-related decision-making and retention. They do not explore the question which specific aspects and design choices of the game contribute to its effectiveness.

Another line of work from Angafor (2020), Irvine (2005) and Ashley (2022) explore games that teach key principles relevant to critical infrastructure (CI), including incident response, system availability, and supplier relationships [28, 34, 35]. However, such games often lack a comparative dimension and do not systematically analyze how game design choices—such as spatial interaction—impact learning outcomes. Furthermore, many of them are not tailored for experimental setups and therefore cannot be directly reused to compare different game types, such as 2D and 3D implementations, under controlled conditions.

Cano (2016) and Prümmer (2021) also attempt to measure how gamified cybersecurity training affects behavior in practice [36, 37], or the study of Angafor (2020) shows how different genres influence user engagement and motivation [28]. Wu (2021) examine specific structural gamification elements, such as leaderboards and challenges, to improve general information security awareness [38]. Yet, none of these studies specifically isolate the spatial dimension as an experimental factor.

While these studies offer valuable insights into the development of effective cybersecurity games, they do not investigate whether the spatial characteristics of a game (such as 2D vs. 3D environments) influence learning outcomes. Only few studies address the spatial dimension explicitly, and mostly in unrelated domains. For example, Zaharias (2017) [17] finds that 2D games may outperform 3D in knowledge transfer, while Aksoy (2019) [20] show improved performance in 3D or VR-based settings due to increased immersion. Kaimara

(2019) reports no significant difference between formats [24], or notes that while 3D environments offer higher situational presence, Mayer (2024) and Parong (2018) states that they do not inherently scale to superior conceptual learning due to increased cognitive load [33, 39]. Consequently, results highly dependent on domain-specific factors.

Most importantly, no existing work directly compares 2D and 3D educational games in the context of cybersecurity training. In many comparative studies, the games differ not only in visual dimension but also in content or mechanics [23, 33], making interpretation difficult. Studies that keep the gameplay consistent often apply only minimal 3D adaptations [18], limiting the potential benefits of spatial interaction. In order to investigate the different environments for educational games in the field of cybersecurity in isolation, we have developed games that are aligned in terms of content and concept.

2. Methodology

In this section we explain our methodology: The games and their learning objectives, the research questions we want to answer, and the survey that we conducted to investigate those questions.

2.1 The Games

We give a brief description of both games and state the desired learning outcomes. Both games should be designed to meet the following learning objectives in cybersecurity with a focus on CI:

1. **Basic understanding of critical infrastructures (CI) and their cybersecurity challenges:** After playing the game, learners are expected to name specific cybersecurity risks and understand how these risks affect the availability of CI services.
2. **Recognition and application of cybersecurity protection measures for CI:** Learners acquire knowledge of basic protection mechanisms, such as intrusion detection systems or encryption, and apply them appropriately in game scenarios.
3. **Detection of cybersecurity threats and incident response in CI:** After playing the game, participants are expected to recognize various cybersecurity threats and suggest appropriate countermeasures. They are also expected to understand qualitative differences between the measures, particularly regarding response time.
4. **Understanding of emergency plans for CI:** Learners should become familiar with contingency plans for responding to cyberattacks on critical infrastructures and demonstrate the ability to apply them in simulated situations.

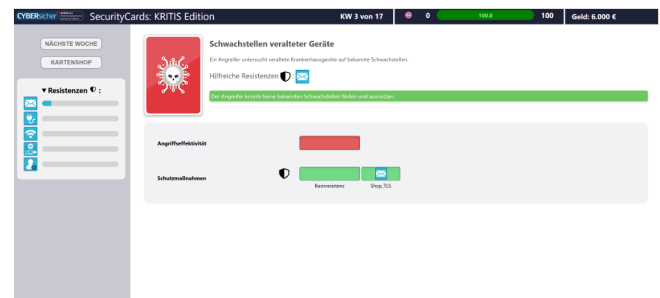
5. Promotion of problem-solving thinking in the context of CI cybersecurity: The game aimed to foster the learners' ability to develop effective solutions to cybersecurity challenges in a timely and efficient manner.

The basic idea is to define a game schematic and instantiate it in two ways that differ only in the presentation of the game. The content should be as identical as possible in order to be able to attribute differences in the learning success of the learners to the spatial dimension of the games. The chief difference between both games is that one is a first-person game situated in a 3D virtual world, and the other is a 2D UI-only game. Yet the game-mechanics, which dictate how a player may interact with the game, are very similar in both cases.

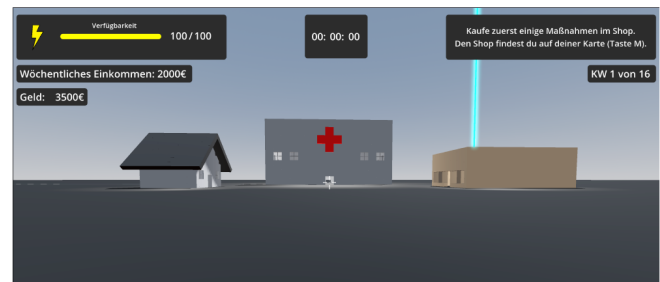
In both games, the goal is to maintain a high availability of the critical infrastructure (in this case, a hospital), while avoiding bankruptcy. The game is lost, if availability falls below a critical threshold (because of cyber incidents), or if the available budget is depleted (because of incidents or excessive spending). Player invest in countermeasures (e.g. redundant internet connection), subject to the available budget. During each turn, random events occur. Some are harmless, some are dangerous and may trigger an incident, if the protective measures prove to be insufficient. For example, investing in employee phishing training will decrease the chance of phishing attack turning into an incident (i.e., the attack was successful). Once an incident occurs, the player must decide how to respond. The available choices again depend on previous purchases (e.g. restoring from backups is only possible if backups were purchased beforehand). Incidents have an effectiveness that may or may not exceed the available protections plus some "base resistance" to model that even unprotected business may be lucky, and even well-protected businesses may be compromised. Both games visualize this using two bars and comparing their length (cf. 1a).

Both games are online games that can be played in a browser. We remark that both games are in German and reference German laws and agencies. This is because the 2D-Game is based on the game "SecurityCards: KRITIS Edition" (KRITIS is the German acronym for critical infrastructure), which was created as part of the project 'BAKSecure'¹. Some adjustments were made to the content to ensure that both games cover the same topics and events. We now describe both games, focusing on the distinguishing factors:

2D-Game The first game is structured as a card-game. A screenshot of the game is shown in fig. 1a. There is a card shop where players purchase protective effects



(a) 2D card game. UI-only [40].



(b) 3D open-world game with multiple locations [41].

Figure 1. Screenshots of the two game instances with different spatial properties

or the option to invoke reactions (like the backups mentioned before). Availability and budget are shown in the top bar. Events are prominently featured in the top third of the screen. On the left, there is an overview of active protective effects (e.g., recent employee awareness training). There is an interactive tutorial level at the beginning. Once that is completed, no further assistance is provided.

3D-Game A screenshot of the 3D-Game is shown in fig. 1b. Various locations were created as an adaptation to the spatial dimension. The player must visit those locations in order to perform certain actions (e.g. a "war room" inside the hospital to make decisions, a store where cybersecurity measures can be purchased, etc.). There are different ways in which the player must react to events. For example, she has to visit the "BSI" (the German federal agency responsible for cybersecurity) to report breaches. In addition, there is a time component: If an event causes damage to the company, a timer counts how long it takes the player to reach the relevant location and carry out a reaction. Longer response times have a negative impact on the availability of the fictional hospital. A tutorial guides the player through the first round. But, as is usual in virtual games, visual effects aid the player in identifying the next location or next action (the column of light in fig. 1b).

Comparison: The 2D-Game consists exclusively of UI elements that offer the player limited options for action and provide a clear structure for the gameplay. In the

¹ See also the acknowledgement on page 16.

Table 1. Comparison of 2D and 3D versions of the Security Cards game

Feature	2D-Game	3D-Game
2D/3D	2D	3D
Time	Turn-based	Real-time with turn-based progression
Space	Discrete	Continuous
Gameplay	Fixed set of actions	Continuous movements and discrete interaction possibilities
World	UI only	Open world
Perspective	Top-down/Tabletop	First-person

3D-Game, on the other hand, players can move freely in a virtual environment. Although the interaction possibilities are limited in the same way as in the 2D-Game, the additional freedom of movement has a considerable influence on the gameplay. In addition, the 3D-Game introduces a temporal component: Movement takes time. This could have an impact on the gaming experience and the evaluation results, which is why these aspects must be taken into account when analyzing the data. A side by side comparison is given in table 1.

We summarize that the timer measuring speed of response and the anchoring of actions to locations in 3D space are the main difference between the two games. In a card game, time and physical localities have no natural correspondence.

2.2 Study Design

We used a between-subjects experimental design to compare two spatial variants of the same serious game (2D UI vs. 3D first-person). Participants were randomly assigned to one of the two conditions (2D: $n = 77$; 3D: $n = 82$). Each participant completed a knowledge pre-test, played their assigned game once, completed an immediate post-test and an EGameFlow questionnaire (measuring immersion and motivation), and then completed a delayed knowledge post-test one week later. Primary outcomes were knowledge gain (pre $\rightarrow$ post), EGameFlow scores, and retention at one week. Statistical analyses included paired tests for within-subject pre $\rightarrow$ post change, Welch's t-tests for between-group comparisons, and Cohen's d as an effect-size measure. Missing data were handled via complete-case analysis.

2.3 Learning Content and Objectives

The game content was developed from typical organisational cybersecurity issues (DDoS, ransomware, preventive vs. reactive measures, incident reporting obligations, cost prioritisation, and availability as a protection

goal). These topics were translated into equivalent tasks and scenario items in both the 2D and 3D versions. The 3D version is realised as a first-person open-world hospital environment and was implemented using the open-source Godot engine [42], while the 2D version presents the same content in a user-interface/card format implemented using React [43]. Learning objectives target mainly the lower to middle levels of Bloom's taxonomy [44, 45] (Remember, Understand, Apply), for example: recall key security concepts, explain appropriate countermeasures, and select correct actions in scenario-based tasks. Higher-order cognitive goals (Evaluate, Create) were not primary aims for this awareness/training intervention.

2.4 Development Criteria and References

The game scenarios and quiz items were developed according to explicit practical and regulatory criteria to ensure relevance for organisational cybersecurity awareness and training. Key development criteria were: (1) relevance to common organisational incidents (DDoS, ransomware, phishing), (2) alignment with employee responsibilities and reporting obligations, and (3) applicability to critical-infrastructure operators and corporate contexts. To ground the scenarios we referenced national guidance and regulations, notably BSI guidance on KRITIS obligations and the German KRITIS strategy, as well as material on NIS2 implementation and sector guidance (see [46–48]). Each quiz item was explicitly mapped to one or more source documents (mapping available in the supplementary material) so that, for example, items on incident reporting reflect BSI requirements while preventive/reactive classification items follow standard incident-response recommendations.

2.5 Research Questions

We study the following research questions:

RQ1 How effective is a 3D virtual, continuous-time game compared to a UI-only card game in improving the understanding of security concepts in critical infrastructures, particularly regarding availability as a key security goal?

RQ2 How does a 3D virtual, continuous-time game influence the engagement and motivation of learners, if compared with a UI-only, turn based card game.

RQ1 There are advantages and disadvantages to using a 3D game for teaching. Although such games are well suited to model real-world environments in detail, a detailed model is not always necessary and might be misleading. All games usually contain abstractions.

For instance, rebooting a server could be modeled as entering a room and clicking a 3D model of a server. This is an oversimplification that will only be apparent to experts — in reality, this process involves logging into the server, notifying clients of downtime etc. It is not usually necessary to interact with the hardware at all. In a 2D card game, the same action could be modeled using a card that reads “reboot server”. Here, the abstraction is obvious, even to non-experts. On the other hand, some relevant aspects, such as the time required to reboot a server before it is again fully operational, cannot be easily modeled easily in a UI-only card-game.

In the context of critical infrastructures, availability is a key security goal. Downtime causes significant damage, and timely responsive action is often required to prevent cascading effects. The perception of time and an accurate model of the physical world would appear to be very important towards internalizing the concept of downtime. With this research question we ask: Is the added detail beneficial for players in achieving their learning objectives in the context of security in critical infrastructures?

RQ2 A game should be engaging and motivate players to strive for better performance and learning outcomes. It seems natural to assume that a 3D game, as described above, is be more engaging. But the virtual model of reality can also conflict with what a player expects based on personal experience regarding distances or facility layout. This may hamper engagement. A 2D card-game is an abstract model of the real-world and cannot conflict with personal experience in the same way. RQ2 asks which of the two kinds of games is more engaging and motivating.

2.6 Survey Structure

The participants were recruited via the “Prolific” platform on a first-come, first-served basis. This platform offers the possibility of recruiting a specific target group to participate in a study for a fee, so there was a monetary incentive [49]. In particular, all participants are aware of the purpose of their participation and have given explicit consent. For this study, only people who reside in Germany, speak German as their native language, and have a professional job (either full-time or part-time) were selected. It was also ensured that each subject was assigned to only one of the two groups. Participation in both groups was excluded, as otherwise it would not be possible to clearly determine the influence of the individual games on the test subjects. The surveys were conducted using the “Tivian” platform [50]. Participants were free to answer only a subset of questions, at their discretion. The study did not use

blinded sampling, as participants were aware that they were taking part in a research study.

Data collection was conducted in compliance with the General Data Protection Regulation (GDPR). Informed consent was obtained from all participants before the start of the study. No unnecessary directly identifying personal data were collected, and the data were processed and stored in pseudonymized form for research purposes only. Participation was voluntary, and participants could withdraw at any time.

Ethical Considerations

All participants were consenting adults, residing in Germany and not part of any vulnerable population. The intervention (playing a video game and completing online surveys) is non-invasive and does not pose any foreseeable physical or psychological risks to participants. No deception was used regarding the purpose, goals, or methods of the study. Accordingly, the ethical implications of this study are well understood, managed, and below the threshold of formal ethics-board review.

The study was structured as follows:

1. **Group division:** The participants are divided into two groups. One group played *2D-Game* ($n = 77$), the other *3D-Game* ($n = 82$).
2. **First survey (pre-test, game-play, GameFlow questions):** The first survey has five steps:
 - i) Demographic questions.
 - ii) A knowledge pre-test determine the participants’ level of knowledge in the area of cybersecurity in critical infrastructures (details below).
 - iii) Gameplay: Participants play the *2D-Game* or *3D-Game*, depending on group.
 - iv) Immediately after the game: A control question to ensure that they have actually played the game.
 - v) Immediately after the game: A survey on motivation and engagement during the game.
 - vi) Finally, participants had the opportunity to leave additional comments.
3. **Second survey (post-test):** One week after completing the first survey, participants receive a second survey, which consists exclusively of the post-test (details below).
4. **Evaluation:** After the second survey is completed, the results of the study are analyzed.

2.7 Pre- and Post-Test-Questions

As the games are in German and the target group is German-speaking people, the surveys were also in German. We provide a translation of the survey below.

The pre-test consisted of 6 tasks:

T 1. “Estimate the cost of the following measures and rank them accordingly.”

The measures, from cheap to expensive, are: 1. Use of an encrypted VPN 2. One-off e-learning training for employees 3. Biometric access controls and use of smart cards 4. Provision of emergency power generators

T 2. “Classify the following measures as preventive or reactive.”

Options: “Use of emergency power generators in the event of a power failure” (reactive), “Use of an alternative transport route in the event of traffic disruptions” (reactive), “Implementation of regular updates to fix known vulnerabilities” (preventive), “Use of an encrypted virtual private network (VPN)” (preventive), and “Multi-factor authentication” (preventive).

T 3. “What protective measures are useful in the event of a DDoS attack?”

Options: (multiple correct): “Use a firewall” (✓), “Use an intrusion detection system” (✓), “Carry out regular backups” (✗), “Use multi-factor authentication and continuous authentication” (✗), and “Set up redundant network structures” (✓).

T 4. “Where must operators of critical infrastructures report a cyberattack?”

Options: “To the Federal Office of Civil Protection and Disaster Assistance (BBK)”, “To the State Office of Criminal Investigation (LKA)”, “To the Federal Ministry of the Interior”, “To the Federal Office for Information Security (BSI)” (✓), or “To the police”.

T 5. “Please estimate how long the following actions usually take”

Options: detect a cyberattack, respond to a cyberattack, report an incident, restore a compromised IT system, implement security measures. The scale was: “a few minutes”, “a few hours”, “1-2 days”, “several days to weeks”, and “several weeks to months”. There is no correct solution; the focus is on how respondents’ assessments change after the games.

T 6. “Which security objective is most affected when it takes a long time to react to an incident?”

Options: “confidentiality”, “integrity”, and “availability” (✓).

The post-test only differs from the pre-test in the following tasks 1 to 3:

T 1. This time the choices were (cheap to expensive):

1. Implementation of regular updates 2. Establishment of satellite communication 3. Establishment of redundant network structures 4. Implementation of regular risk analyses and penetration tests

T 2. This time the options were “Conduct regular risk analyses and penetration tests” (preventive), “Use an intrusion detection system to detect ongoing attacks” (reactive), “Encrypt all data transmissions with TLS” (preventive), “Use access controls so that only authorized personnel have access to critical areas” (preventive), and “Use of satellite communication in the event of a failure of regular communication channels” (reactive).

T 3. “What protective measures are useful in the event of a ransomware attack?”

Options: “Use an intrusion detection system” (✓), “Use a virtual private network (VPN)” (✗), “Deploy an incident response team” (✓), “Encryption with TLS” (✗), and “Perform regular backups” (✓).

Rationale T1 to T3 cover knowledge that is implicitly taught in the game: Players have exposure to these facts, but do not have to perform these specific tasks themselves. In the event of an attack, both games offer only suitable countermeasures, for example. These tasks are designed to assess if implicit knowledge improves as a result of the games. Regarding T4, players have to know where to report incidents in order to complete the 3D-Game. But in the 2D-Game, there is no requirement to report incidents. This task is designed to check that explicitly taught tasks have an impact. Task 5 covers the influence of the time dimension on the perception and expectations of players regarding realistic response times. Finally, T6 assesses how well the primary security objective ‘availability’ is internalized by players and if the 3D version has a beneficial impact on this.

EGameFlow To measure the motivation and engagement of the participants we used 17 questions from the *EGameFlow* scale [51] including items such as “I had fun while playing the game” or “The game motivates the player to actively use the knowledge imparted,” using a 5-point Likert scale (0 = strongly disagree to 4 = strongly agree). The *EGameFlow* model was developed to systematically measure user enjoyment in e-learning games in order to better understand their learning effectiveness. The scale comprises eight dimensions such as immersion, challenge and knowledge gain. We have selected a subset of those question to focus on our research questions and dropped questions that are irrelevant to that goal. For example, the category *Social Interaction* does not apply to our study, and the category *Challenge* goes beyond what these games implement

(e.g., multiple levels to accommodate more advanced players).

2.8 Sample

Participants were recruited to represent employees and professionals who would use the game for workplace cybersecurity training. The game scenarios and learning objectives were framed for organisational use (employee awareness and decision making in critical-infrastructure or corporate environments).

Population 2D-Game Participants ($n = 77$) came from various sectors: 18.2% of the participants work in the service sector ($n = 14$), making them the most frequently represented group. 14.3% work in cross-sectoral or other areas ("Other", $n = 11$), while 9.0% come from the technology and telecommunications sector ($n = 7$). In the pharmaceutical and healthcare sectors, 7.8% of the participants are represented ($n = 6$), and 6.5% work in the emergency services (e.g., fire department or police, $n = 5$); likewise, 6.5% work in the metal and electronics industry ($n = 5$). 5.1% work in the chemical and raw materials industry ($n = 4$). 3.9% each come from the transport and logistics industry and the energy and environment sector (each $n = 3$). Smaller shares are recorded by sectors such as finance, internet, construction, e-commerce, media and retail. Other individual respondents come from the leisure, society, consumer goods (FMCG), tourism, politics, crafts and other niche sectors.

Most participants work in a large company with more than 500 employees ($n = 34$; 44.2%). 18.2% work in a small company with fewer than 9 employees ($n = 14$), 15.6% in companies with 50 to 249 employees ($n = 12$), 11.7% in companies with 10 to 49 employees ($n = 9$), and 9.1% in companies with 250 to 499 employees ($n = 7$). One participant did not answer this question.

With regard to the highest educational qualification, 2.4% of participants ($n = 2$) have a secondary school leaving certificate (intermediate school leaving certificate), 3.7% ($n = 3$) have a doctorate, and 6.1% ($n = 5$) each hold a master craftsman's, business administrator's, or specialist management qualification, or a degree from a university of applied sciences. 11.0% ($n = 9$) have completed general vocational training. 18.3% ($n = 15$) hold an Abitur, 19.5% ($n = 16$) have a Master's or Magister degree, and 26.8% ($n = 22$) have a Bachelor's or Diplom university degree.

Regarding the departments in which the participants work, the largest group (22.0%, $n = 18$) is employed in IT and EDP. 13.4% ($n = 11$) work in research and development. 6.1% of participants each work in education and training, sales, and in the management of smaller companies (up to 249 employees, $n = 5$). 4.9% ($n = 4$) work in administration and in design, music, dance or drama departments. 3.7% ($n = 3$) are employed

in traffic, logistics and transportation, and 2.4% ($n = 2$) each in the areas of crafts, production or marketing. A further 12.2% ($n = 10$) work in individual, non-categorized areas such as public safety, legal services or editing.

The majority of respondents (51.2%, $n = 42$) stated that there was no known security incident in their environment. In 20.7% of cases ($n = 17$), an incident was recorded, but the participants were not directly involved. 12.2% ($n = 10$) were actively involved in resolving an incident. 9.8% ($n = 8$) reported having been affected by an incident themselves, while 6.1% ($n = 5$) stated that they had been the direct or indirect cause of an incident or the target of an attack – for example, through phishing.

When asked whether the participants have responsibility beyond their personal area, 70.1% responded that they have no more responsibility than for their personal area ($n = 54$), while 29.9% reported that they also make decisions with an impact on IT security in the company, but have other important tasks as well.

Regarding gaming habits: 29.0% of participants play video games several times a week ($n = 22$), 25.0% play daily ($n = 19$), 22.0% ($n = 17$) play less than once a week, 12.0% never play video games ($n = 9$) and another 12.0% play once a week ($n = 9$).

Population 3D-Game The participants ($n = 82$) in the survey come from different sectors. The largest proportion came from the service sector (17.1%, $n = 14$), followed by the "Other" category (15.9%, $n = 13$) and the technology and telecommunications sector (13.4%, $n = 11$). Other notable groups include the metal and electronics industry and the pharmaceutical and healthcare sectors (8.5% each, $n = 7$). Smaller proportions can be found in the construction, retail, internet, media, and finance and insurance sectors, among others.

46.3% of the participants work in a large company with more than 500 employees ($n = 38$), 19.5% work in companies with 50–249 employees ($n = 16$), 13.4% each work in a micro-company with fewer than 9 employees ($n = 11$) and in a small company with 10–49 employees ($n = 11$). 7.3% are employed in companies with 250–499 employees ($n = 6$).

The participants' educational qualifications are distributed as follows: 29.3% hold a Master's degree ($n = 24$), 22.0% have a high school diploma ($n = 18$), 18.3% hold a Bachelor's or Diplom degree ($n = 15$), 12.2% have completed vocational training ($n = 10$), 7.3% have a university of applied sciences degree ($n = 6$), 3.7% hold an intermediate school leaving certificate ($n = 3$), and 2.4% each have a master craftsman/business administration degree ($n = 2$) or a basic/main school leaving certificate ($n = 2$). 1.2% each have a doctorate or a degree from a vocational college ($n = 1$ respectively).

More than a third of the respondents (41.5%) stated that they were not aware of any IT security incidents. 26.8% reported an incident in which they were not directly involved. A further 12.2% had been actively involved in resolving an incident, 7.3% had been affected themselves, and 8.5% stated that they had been the target of an attack or the (co-)cause of an incident.

The analysis shows that the vast majority of respondents (67.1%, $n = 55$) have no responsibility for IT security beyond their own area. Around a third (31.7%, $n = 26$) make security-related decisions, even if this is not their primary task. Only one person (1.2%, $n = 1$) is primarily responsible for IT security.

The participants work in a wide range of departments. The largest group works in IT and EDP (24.4%), followed by "Other" (11.0%), research and development (9.8%), and production (8.5%). Other frequently mentioned areas include administration, sales, media, and customer service.

35.4% of participants play video games several times a week ($n = 29$); another 35.4% play video games daily ($n = 29$). 19.5% play less than once a week ($n = 16$), 7.3% play once a week ($n = 6$), and 2.4% never play video games ($n = 2$).

Table 2 provides an overview of some key characteristics of the sample.

Table 2. Sample characteristics by condition

Characteristic	2D ($n=77$)	3D ($n=82$)
Sector (largest)		
Service	14 (18.2%)	14 (17.1%)
Other / Cross-sector	11 (14.3%)	13 (15.9%)
Tech & Telecom	7 (9.0%)	11 (13.4%)
Pharma & Healthcare	6 (7.8%)	7 (8.5%)
Metal & Electronics	5 (6.5%)	7 (8.5%)
Company size		
<9 employees	14 (18.2%)	11 (13.4%)
10–49	9 (11.7%)	11 (13.4%)
50–249	12 (15.6%)	16 (19.5%)
>500	34 (44.2%)	38 (46.3%)
Education (selected)		
Master	16 (19.5%)	24 (29.3%)
Bachelor / Diplom	22 (26.8%)	15 (18.3%)
Department (selected)		
IT / EDP	18 (22.0%)	20 (24.4%)
R&D	11 (13.4%)	8 (9.8%)
Incident experience (no known)	42 (51.2%)	34 (41.5%)
Responsibility for IT security (no)	54 (70.1%)	55 (67.1%)
Gaming frequency (daily/weekly)	41 (53.0%)	58 (70.8%)

Sample characteristics and group comparability

Overall, the two samples for the 2D-Game ($n = 77$) and the SecurityCards 3D game ($n = 82$) show a broadly comparable demographic and professional profile. In

both groups, the largest proportion of participants is employed in the service sector, followed by a heterogeneous mix of cross-sectoral or "Other" professions as well as technology and telecommunications. Classical critical-infrastructure related domains such as health-care, emergency services, energy and transport are represented in both samples, although the distributions across specific sectors differ slightly between the two groups.

With regard to company size, participants from large organisations with more than 500 employees form the largest subgroup in both conditions. At the same time, a substantial share of respondents in each group works in micro- and small enterprises. This distribution reflects the fact that cybersecurity and incident response are relevant for organisations of very different sizes and resource levels. The educational level in both samples is comparatively high, with the majority of participants holding at least a higher secondary degree and a large proportion reporting a Bachelor's, Master's or equivalent university degree. Because the participants are generally well educated, it is likely that they could follow the cybersecurity content and handle the game-based learning approach without major difficulties.

In terms of professional roles, IT and EDP form the largest departmental group in both samples, followed by research and development and a wide range of further areas such as administration, education and training, production, media or customer service. Thus, the games were not only evaluated by IT specialists, but also by employees from non-technical departments, which aligns with the goal of using serious games to raise awareness among a broad professional audience.

Prior exposure to IT security incidents and the degree of security-related responsibility are also comparable across the two conditions. In both samples, around half of the participants report no known incident in their environment, while a notable fraction has either been directly involved in resolving an incident or has been personally affected by an attack. Approximately one third of respondents in each group indicate that they make decisions with an impact on IT security beyond their own work area, even though this is not their primary task. Only very few participants report being primarily responsible for IT security.

Finally, the gaming habits of both groups indicate a high level of familiarity with digital games. In the 2D-Game sample, more than half of the participants play video games at least once a week, and in the 3D-Game sample around seventy percent report playing several times a week or daily. Only a small minority in either group never plays video games. This suggests that most participants are accustomed to basic game mechanics, user interfaces and navigation in virtual environments. As a result, usability issues are unlikely to dominate their experience, and differences between the 2D and

3D conditions can be interpreted primarily in terms of game design rather than basic accessibility.

Taken together, the two samples are similar enough in their demographic composition, professional background, prior security experience and gaming habits to allow for a meaningful comparison of learning outcomes and engagement between the 2D and 3D game versions. At the same time, the described characteristics need to be taken into account when discussing the external validity of the findings.

3. Results

We now describe the results of the survey. First, the quantitative results and then the qualitative results.

3.1 Quantitative Results

We explain how questions were scored and present a comparison of means between the pre- and post-test scores per group (2D group and 3D group). These tests are always paired t-tests. Moreover, we show comparisons of the pre- and post-test means between the groups. These are always two-sided unpaired Welch's t-test scores. Our significance level is $\alpha = 0.05$. All statistical analyses were performed in a Jupyter/Colab notebook using Python (Python 3.12) (pandas, numpy, scipy); inferential tests (paired t-tests, Welch's two-sample t-tests, and nonparametric alternatives where applicable) and Cohen's d effect-size calculations were implemented in the notebook [52–57].

Task 1. In the first task, certain countermeasures had to be sorted according to their financial costs. To score this question, we counted how many are inserted in their correct position. The same question was asked in pre- and post-test, but with a different set of countermeasures to be sorted.

The cost of security measures plays a role in both games, players need to protect the hospital using a constrained budget. However, ranking measures solely based on their technical description is not part of the game: Players always have access to information about the cost of countermeasures.

In the 3D-game, the participants performed statistically significantly better in the post-test compared to the pre-test (cf. table 3, $p=0.007$, Cohen's $d = 0.32$, indicating a small to medium effect size. The improvement in the 2D group is not statistically significant. It should be noted, however, that the p value is low enough to suggest that further investigation is needed. In contrast, there are no statistically significant differences between 2D and 3D when comparing the average pre- and post-test scores. Note that for the pre-test, we would not expect a difference, unless the two groups suffer from some form of selection bias. However, the comparison

Table 3. Comparison of relative estimated cost of countermeasures

(a) Pre vs. Post test per group.

Group	Mean		p
	Pre	Post	
2D	0.301	0.493	0.080
3D	0.256	0.590	0.007

(b) 2D vs. 3D comparison.

	Mean		p
	2D	3D	
Pre	0.301	0.256	0.642
Post	0.493	0.590	0.426

Table 4. Comparison of classification of countermeasures.

(a) Pre vs. Post test per group.

Group	Mean		p
	Pre	Post	
2D	4.425	4.493	0.504
3D	4.269	4.205	0.591

(b) 2D vs. 3D comparison.

	Mean		p
	2D	3D	
Pre	4.425	4.269	0.268
Post	4.493	4.205	0.026

of post-test means also shows no statistically significant difference. This indicates that the effect is modest overall. Further research is indicated.

Task 2. The second task was to classify five countermeasures as either preventive or reactive. Again the set of countermeasures differed between pre-and post-test. To score this question, we again counted how many are inserted in their correct position.

Neither game places explicit emphasis on this classification. Whether a countermeasure is reactive or preventive in nature only affects *when* the game allows selecting it (after an incident or before).

There is no statistically significant difference in the pre-test and post-test scores (cf. table 4). While both groups show no statistically significant difference in pre-test scores (which is expected, as discussed above), we do note a statistically significant difference in post-test scores between the 2D and 3D groups ($p = 0.026$). This is surprising, particularly since the paired t-test shows no statistically significant difference between pre- and post-test for the 3D group. However, the effect size for the post-test comparison is very

Table 5. Comparison of selecting effective countermeasures

(a) Pre vs. Post test per group.			
Group	Mean		<i>p</i>
	Pre	Post	
2D	0.178	0.205	0.567
3D	0.192	0.256	0.320
(b) 2D vs. 3D comparison.			
	Mean		<i>p</i>
	2D	3D	
Pre	0.178	0.192	0.823
Post	0.205	0.256	0.460

small (Cohen's $d = -0.12$), indicating that, despite statistical significance, the practical difference between the groups is negligible.

It seems likely that this inconsistent result is due to the different test-statistics that are computed for a paired t-test and an unpaired Welch's t-test. Note also that $p = 0.026$ is still relatively close to our confidence level of 0.05. We cannot place much confidence in these findings. As a result, our experiment cannot support an effect (positive or negative) on the ability to classify countermeasures as responsive or preventive and cannot support the existence of any effect of the additional design choices in the 3D-Game, in particular.

Task 3. Participants classified what countermeasures are effective in a given scenario ("DDoS" in the pre-test; "Ransomware" in the post-test). Here we only give one point (True/False) per answer option. The results are shown in see table 5. We see no statistically significant improvement in the test scores for either of the two games. The comparison of pre-test and post-test means between both groups also shows no statistically significant difference.

Both games do teach selecting appropriate countermeasures. However, the player chooses only from a set of pre-selected options, all of which are generally appropriate, but differ in, e.g., their cost or robustness. So again, the games teach this knowledge in passing, but stop short of explicitly making it part of the gameplay.

Task 4. Participants had to select the appropriate agency to which CI operators need to report cyberattacks. The question and correct answer were identical in the pre- and post-test for both games. To evaluate the question, the correct answer was scored with one point and the incorrect answers with zero points.

The information needed to answer this question is taught exclusively and explicitly in 3D-Game. The group playing 2D-Game therefore serves as a control group for this task, so that it can be observed to

Table 6. Comparison of correctly reporting cyberattacks

(a) Pre vs. Post test per group.			
Group	Mean		<i>p</i>
	Pre	Post	
2D	0.740	0.767	0.531
3D	0.821	0.961	< 0.001
(b) 2D vs. 3D comparison.			
	Mean		<i>p</i>
	2D	3D	
Pre	0.740	0.821	0.235
Post	0.767	0.962	< 0.001

what extent the answers change in the post-test if the necessary knowledge has not been taught.

As can be seen in table 6, there were significant improvements for the 3D group. This is reflected in both the scores of the paired t-test for the 3D group (Cohen's $d = 0.40$, moderate effect size) and the two-sample t-test between the 2D and 3D groups applied to the post-test scores (Cohen's $d = -0.60$, moderate effect size). Note that, as expected, there was no statistically significant difference between both groups prior to playing the games.

Task 5. In the fifth task, participants estimated the time certain countermeasures and responses take to implement. Possible responses and their numerical scores are as follows: 'Several weeks to months': 5, 'Several days to weeks': 4, '1-2 days': 3, 'Few hours': 2, 'Few minutes': 1.

The 2D-Game makes no reference to the time required to implement measures or to respond to incidents. The 3D-Game has an implicit temporal component: Players need to move to specific locations to initiate responses and these places have different distances to the hospital complex.

There was no statistically significant difference between the 2D and 3D groups, either before or after playing the game. Furthermore, no significant change was observed from pre- to post-test within the 2D group (see table 7a). However, within the 3D group, there was a statistically significant reduction in estimated implementation times from pre- to post-test ($p = 0.045$, Cohen's $d = -0.23$), indicating a small effect size.

Task 6. Participants were asked which primary security objective is most negatively impacted, if responding to an incident takes a long time. Choices were 'Availability' (correct), 'Integrity', and 'Confidentiality'. The correct answer was scored with one point and incorrect answers with zero points.

Table 7. Comparison of estimated duration

(a) Pre vs. Post test per group.

Group	Mean		<i>p</i>
	Pre	Post	
2D	4.342	4.247	0.321
3D	4.167	3.910	0.045

(b) 2D vs. 3D comparison.

	Mean		<i>p</i>
	2D	3D	
Pre	4.342	4.167	0.223
Post	4.247	3.910	0.058

Table 8. Comparison of determining most affected security objective

(a) Pre vs. Post test per group.

Group	Mean		<i>p</i>
	Pre	Post	
2D	0.520	0.658	0.017
3D	0.487	0.641	0.013

(b) 2D vs. 3D comparison.

	Mean		<i>p</i>
	2D	3D	
Pre	0.520	0.487	0.684
Post	0.657	0.641	0.833

3D-Game illustrates the relationship between availability and response time particularly well, as the additional dimension enables an intuitive representation of time sequences. With 2D-Game, on the other hand, this connection is less obvious and only reflected in a numeric availability score.

Both the 2D and 3D groups showed a statistically significant improvement from pre- to post-test in the protection goal classification task (2D: $p = 0.017$, Cohen's $d = 0.29$; 3D: $p = 0.013$, Cohen's $d = 0.29$; see table 8). The effect sizes indicate a small to medium improvement in both groups, with the effect being slightly greater in the 3D group. However, there were no statistically significant differences between the two groups in either the pre-test or the post-test means.

EGameFlow. The results for the EGameFlow questions are shown in table 9. We used a 5-point Likert scale (from 0 = strongly disagree to 4 = strongly agree). For each item, we report mean values for both groups, the statistical significance (p -value), and the effect size (Cohen's d). The p -value indicates whether the observed difference is statistically significant ($p < 0.05$), while

Table 9. Comparison EGameflow 2D and 3D game versions (incl. Cohen's d)

Item	Mean		<i>p</i>	Cohen's d
	2D	3D		
1. I had fun playing the game.	2.616	2.494	0.493	0.17
2. The game has my full attention.	3.699	3.545	0.135	0.19
3. Most of the activities in the game contribute to learning.	2.306	2.500	0.258	-0.27
4. I am burdened with tasks that seem irrelevant.	1.438	1.538	0.527	-0.12
5. The tasks in the game are overwhelming.	1.507	1.064	0.012	0.37
6. The general objectives of the game were clearly presented.	2.932	3.000	0.671	-0.08
7. Intermediate objectives were clearly presented.	2.397	2.821	0.008	-0.46
8. I understand the learning content of the game.	2.708	2.974	0.123	-0.30
9. While playing the game, I forget how time passes.	2.671	2.551	0.535	0.03
10. While playing the game, I am no longer aware of my surroundings.	2.233	2.132	0.568	0.09
11. While playing the game, I temporarily forget my everyday worries.	2.247	2.192	0.768	0.05
12. I can put myself in the game's shoes.	2.736	2.736	1.000	-0.04
13. The game expands my knowledge.	2.486	2.692	0.223	-0.26
14. The game motivates the player to actively use the knowledge imparted.	2.408	2.538	0.472	-0.23
15. I would like to learn more about the knowledge imparted.	2.347	2.378	0.889	-0.09
16. The game promotes my problem-solving skills in relation to cybersecurity.	2.041	2.385	0.070	-0.36
17. The game gave me a better understanding of the challenges of cybersecurity.	2.630	2.769	0.437	-0.17

Cohen's d quantifies the magnitude of the difference (with values of 0.2, 0.5, and 0.8 commonly interpreted

as small, medium, and large effects, respectively [53]). A positive value for Cohen's d indicates a higher mean for the 2D group, while a negative value indicates a higher mean for the 3D group.

Two-sample t -tests for question 5 ("The tasks in the game are overwhelming") and question 7 ("Intermediate objectives were clearly presented") revealed significant differences between groups. Participants who played the 2D game were more likely to agree that the tasks in the game were overwhelming compared to those who played the 3D game ($p = 0.012$, $d = 0.37$), indicating a small-to-moderate effect size. Regarding question 7, players of the 3D game expressed stronger agreement that intermediate objectives were clearly presented ($p = 0.008$, $d = -0.46$), representing a moderate effect size. For all other items, differences between groups were neither statistically significant nor practically relevant (all $p > 0.05$, $d < 0.3$).

Reporting both p -values and effect sizes provides a more comprehensive understanding of the group differences, highlighting not only statistically significant results but also their practical relevance.

3.2 Qualitative Results: Open-ended Responses

To complement the quantitative results with deeper insights into participants' subjective experiences, the open-ended responses to the question "Is there anything else you would like to tell us about the game and the survey and were there any technical issues in the game?" were subjected to a qualitative content analysis. The evaluation followed a thematic coding approach: All responses were carefully reviewed, and recurring topics and patterns were identified. Based on these observations, the answers were grouped into content categories (codes) to highlight the main themes and notable aspects. The main coding categories that emerged were:

- Praise and positive feedback,
- Suggestions for improvement,
- Technical comments/issues and
- Other remarks.

Each response was assigned to one or more of these categories according to its content. This method allows for both the identification of common themes and the structured presentation of unique individual perspectives.

Qualitative Analysis: 2D Game. A total of 24 of the 77 participants provided open-ended feedback regarding the 2D version of the game. The responses revealed several recurring themes, which were grouped as follows:

- **Positive Feedback ($n=8$):** Many participants described the 2D game as "fun", "entertaining", or "well-made". Several users appreciated the educational aspect and the overall concept, indicating that the game was engaging and helpful for learning purposes.
- **Suggestions for Improvement ($n=10$):** A large portion of the comments included constructive suggestions. The most common requests were for additional levels or longer gameplay ("Could be longer", "More levels would be nice"). Multiple participants mentioned a need for clearer instructions or tutorials at the beginning ("More explanation at the start would help"). The control scheme was also frequently noted, with some finding it "unintuitive".
- **Technical Issues ($n=2$):** Only two participants experienced technical difficulties (for example, one clicked the 'Back' button, meaning they had to restart the game).
- **Other Comments ($n=4$):** A limited number of responses included neutral observations.

Qualitative Analysis: 3D Game. 34 of 82 Participants also provided open feedback for the 3D version. The following themes emerged:

- **Positive Feedback ($n=11$):** Eleven participants described the 3D game as "fun", "good" or "interesting". The graphical presentation and immersive atmosphere were specifically highlighted by several players.
- **Suggestions for Improvement ($n=8$):** Six participants wished for better instructions or tutorials ("No instructions," "More hints at the beginning needed"). Two participants found the game too challenging or had trouble with orientation and navigation.
- **Technical Issues ($n=16$):** Sixteen participants reported technical and performance issues, such as lag, bugs, or graphical glitches. The game's performance on less powerful computers was a recurring concern.

Comparison: 2D vs. 3D Game. Participants in both game versions provided open comments that reveal distinct user experiences. While both games were generally well received and described as enjoyable, the 3D version elicited more feedback overall, especially regarding its graphics and immersive atmosphere. However, technical issues were much more common in the 3D game, with several users reporting lag, bugs, and performance problems—particularly on weaker computers. For the 2D game, suggestions focused

more on wanting additional levels, longer gameplay, and clearer instructions; issues with the controls were mentioned but less frequently. In contrast, feedback for the 3D game highlighted difficulties with navigation and orientation, as well as the need for better tutorials. Overall, the 2D game was perceived as more accessible and less stressful, while the 3D game was seen as visually impressive but more demanding and technically challenging.

Supplementary Feedback from Participants Who Played Both Versions. In addition to the main participant groups, three individuals played both the 2D and 3D versions and provided direct comparative feedback. All three stated that the 2D game was noticeably more pleasant and less stressful to play. This suggests that, despite the potentially higher immersion of the 3D environment, ease of use and reduced cognitive load played an important role in the perceived user experience. While the sample size is too small for generalisation, these insights highlight the relevance of intuitive controls and low entry barriers in serious games, and suggest avenues for future research.

Note: The feedback from these three participants, who played both the 2D and 3D versions, was not included in the main quantitative analyses of the individual groups, but is presented here as an additional qualitative perspective.

4. Discussion

The results detailed in section 3 illustrate both the effectiveness, as well as the limits, of our educational games. Because the study targets employee awareness in organisational settings, we discuss applicability for corporate training programs and CI operator in-service training rather than general public education. On the one hand, we see a direct connection between learning success and the content that was explicitly taught in the games. This is highlighted by Task 4 (“Where do critical infrastructure operators have to report a cyberattack?”), which is taught explicitly in the 3D-Game, but not in the 2D-Game. There is a statistically significant gain in the 3D group and a significant difference between the post-test means. Additionally, we do note that the strong focus on availability as a core winning condition of the games (both 2D and 3D) seems to also positively affect the learning success (Task 6). On the other hand, we note that tasks 1, 2, and 3 did not benefit from the games in a clear and consistent way (we will discuss the effects unique to the 3D group further below). The corresponding content was taught only in passing, or requires the capability to transfer knowledge to other domains. This interpretation aligns with the EGameFlow results: only item 5 (overwhelm) and item 7 (clarity of intermediate objectives) differed significantly between versions,

suggesting that guidance and clarity—not spatial realism—drive perceived usability in our setting.

Turning toward the spatial and temporal properties only present in the 3D-Game, we do observe significant effects on the tasks. Players of the 3D-Game showed *lower* estimates for the duration of implementing countermeasures (Task 5) after playing the game. The effect is modest, because it is not also present in post-test mean comparison between 2D and 3D. At first, this seems surprising, given that time plays a major role in the 3D-Game and is, in fact, scored. But the 3D game implements time-lapse primarily as a side-effect of player movements: Moving requires time, but once at the destination, the desired action can be triggered instantly. Since the distances are modest overall, this seems to impart a feeling of efficiency and timeliness on the players.

Players of the 3D-Game also scored better at sorting countermeasures by cost (Task 1). This effect is modest: It is only visible in the pre- and post-mean comparison. There is no clear explanation for this effect based on the structure or content of the game. Costs are relevant in both games, because countermeasures cost money and the constrained budget forces players to take this into account. But this holds true for both games. Neither game makes a unique effort to convey relative costs of countermeasures to the players. Altogether, it is unclear what causes this effect. Since the effect is not very strong, we assume that it is due to an undetected bias in the groups.

Turning to immersion and motivation, it is interesting to note that the 3D-Game is not perceived as better in either of these dimensions. However, it does seem to be better explained. That effect is very likely due to the added assistance: While both games have a tutorial level, the 3D-Game offers continuous hints (column of light and a map of the world). The 3D-Game thus has a more visual tutorial and, unlike the 2D-Game, continues to assist the player throughout the game.

A deeper interpretation of the observed pattern suggests that the spatial and temporal features of the 3D version primarily alter perceptual and situational judgments rather than the underlying cognitive processes required for robust knowledge acquisition. Specifically, the modest decrease in estimated implementation time likely reflects an experiential impression produced by movement and time-lapse mechanics, not a durable change in procedural understanding. Likewise, the isolated improvement in cost-sorting may stem from incidental salience effects in the 3D presentation or residual sample bias, rather than from enhanced conceptual learning. These observations imply a mediating role of attention, perceived efficiency, and guidance: 3D elements change what players notice and how they feel about tasks, but do not automatically induce deeper

elaboration or transfer unless the design explicitly supports such cognitive processes.

Theoretically, our results nuance accounts that equate immersion with superior learning: immersion appears to modulate perceptual and motivational states but does not guarantee improved encoding or transfer. This supports a framework where spatial realism benefits learning only when it scaffolds elaborative processing, provides retrieval-relevant cues, or reduces extraneous cognitive load. In other words, immersion is a potential facilitator, not a sufficient condition, for learning gains.

Compared to prior studies that report mixed outcomes for 2D vs. 3D formats, our findings align with work showing little or no advantage for 3D unless spatial reasoning is central to the task, while contradicting reports that attribute consistent benefits to immersive formats. Differences in prior results likely reflect heterogeneity in domain demands, sample characteristics, and whether the spatial features were tightly coupled to the learning objectives. We therefore place our results in the middle of this spectrum: 3D can change perception and some task-specific responses, but its effect on core learning outcomes depends on how directly the spatial mechanics instantiate the target concepts.

The qualitative findings add depth to the quantitative results by highlighting user experience aspects that may not be fully captured by standardized measures. For example, the repeated mention of intuitive controls and low stress in the 2D game suggests that user-friendliness is a critical factor for engagement and learning.

In summary, while a spatial and temporal component influence players do influence the scores in some tasks, we cannot find evidence that they have significant beneficial effect on the learning success. Similarly they do not appear to boost the immersive and motivational properties of the games. Accordingly, design effort might better prioritise clear goals and ongoing guidance over richer 3D presentation for awareness training. Any added cost incurred by implementing such features should be carefully evaluated.

Practical Implications

Practically, organisations and developers should prioritise instructional alignment and guidance over graphical fidelity when resources are limited. We suggest to use 3D only when the learning objectives require spatial interaction, orientation, or embodied practice; otherwise a well-designed UI-based implementation with clear intermediate objectives and continuous guidance will likely be more cost-effective. If 3D is chosen, invest in onboarding (robust tutorials, navigation aids) and performance optimisation to avoid technical friction that can undermine learning and motivation.

Limitations

As detailed in section 2.8, subjects had a significant amount of gaming experience, potentially tainting the results of our study. It is unclear if and how prior gaming experience would affect the results. It is possible that the effectiveness of learning games is not impacted at all. However, it is plausible to assume that certain aspects of the EGameFlow survey and the perceived usability of the games are influenced by the fact that many participants are already familiar with typical game mechanics, user interfaces, and navigation in digital environments. In particular, the high level of prior gaming experience in both groups may bias the results, as participants are likely more accustomed to game-based interaction than the average employee. Future studies should therefore deliberately include participants with limited or no gaming experience to assess whether the findings generalise to a broader workforce.

In addition, the composition of our samples imposes further limitations on the external validity of the findings. Both groups primarily consist of professionals with medium to high educational qualifications, and a large share of participants holds a Bachelor's, Master's or equivalent university degree. At the same time, a considerable proportion of respondents in each condition works in large organisations with more than 500 employees, while the rest are distributed across micro, small and medium-sized enterprises. This suggests that our results predominantly reflect the perceptions of a relatively well-educated and comparatively technology-affine workforce in organisations that, in many cases, have at least some formalised structures for IT and cybersecurity.

Moreover, the sectoral distribution of the participants is heterogeneous and includes service industries, technology and telecommunications, healthcare, emergency services, and several other domains. While this variety supports the general relevance of serious games for cybersecurity training across different sectors, it also means that some critical-infrastructure domains are only represented by small subsamples. Given the limited ecological validity, the findings should not be over-generalised to all CI operators or to employees with different educational backgrounds and roles, nor to other sectors or less resourced organisations.

The open-ended responses provide valuable insights into the participants' subjective experiences; however, their validity is limited by the voluntary nature of the feedback and the small group size—particularly among those who tested both games. Furthermore, categorizing the responses involves a certain degree of subjectivity.

Another important aspect is the time frame of the retention assessment: the results are based solely on a follow-up test one week later. Consequently, the

findings should be interpreted with caution regarding long-term learning effects, as the study does not capture retention beyond this short time window.

Taken together, these factors indicate that our findings should be interpreted as a first step towards understanding the role of spatial and temporal game elements in cybersecurity training, rather than as definitive evidence that can be directly transferred to all employee populations. Future studies should aim to replicate and extend our results with more diverse samples, including participants with lower levels of formal education, limited gaming experience, and a more targeted coverage of specific critical-infrastructure sectors.

5. Conclusion

We studied the impact of spatial and temporal game elements on learning success in IT security. To this end, we developed two games that share the same basic structure and game mechanics, but differ in their game space. One is a UI-only card game, the other a 3D first-person game situated in a virtual 3D world. We then conducted a study to assess the impact of these different game spaces. Participants were allocated to two independent groups (between-subjects design) and they were quizzed on their IT-security knowledge and were then asked to play either the 3D-Game or the 2D-Game. Subsequently, they answered a questions about the motivational and immersive properties of the game. After one week, they were again quizzed on their IT-security knowledge.

Our findings indicate that the 3D-Game has an impact on the perception of time and seems to be more intuitive. But we could not find evidence that it leads to better learning outcomes than the 2D-Game. This is particularly interesting in the light of the higher development costs of 3D games. Additionally, our study cannot support the hypothesis that 3D games are more engaging or motivating than their 2D counterparts.

In sum, spatial and temporal game mechanics reliably affect players' perceptions and certain task-specific judgements, but they do not by themselves produce superior knowledge gains in our setting. Given the substantially higher development costs of 3D worlds, stakeholders should weigh these perceptual benefits against cost and technical risk and reserve immersive implementations for scenarios where spatial affordances are pedagogically essential.

Future Work We could show that spatial and temporal elements in our 3D-Game have an impact on the perception of how long countermeasures take to implement. But this impact is not aligned with learning outcomes, since participants actually game lower estimates after the game than they did before. In future work, we aim to study how the time and spatial elements of the game must be tweaked to positively

impact learning success. As mentioned, future research should also attempt to reproduce the results found in our present study using a larger and more diverse (in terms of gaming experience) group of participants.

Acknowledgement

This work has been partially developed within the project "BAKSecure" which is funded by the German Federal Ministry for Economic Affairs and Energy. Funding code 01MS23002A. The authors are responsible for the content of this publication.

References

- [1] Lehto M. Cyber-attacks against critical infrastructure. In: *Cyber security: Critical infrastructure protection*. Springer; 2022. p. 3-42.
- [2] Research GV. Cyber Security Training Market Size, Share & Trends Analysis Report; 2023. Accessed: 01-Apr-2025. Available from: <https://www.grandviewresearch.com/industry-analysis/cyber-security-training-market-report>.
- [3] Fund IM. Cyber Risk: A Growing Concern for Macroeconomic Stability. International Monetary Fund; 2024. Accessed: 01-Apr-2025. Available from: <https://www.imf.org/en/Publications/GFSR/Issues/2024/04/11/global-financial-stability-report-april-2024>.
- [4] Tolossa D. Importance of cybersecurity awareness training for employees in business. *Vidya-A Journal of Gujarat University*. 2023;2(2):104-7.
- [5] Shillair R, Esteve-González P, Dutton WH, Creese S, Nagyfejeo E, von Solms B. Cybersecurity education, awareness raising, and training initiatives: National level evidence-based results, challenges, and promise. *Computers & Security*. 2022;119:102756.
- [6] Hadlington L. The "human factor" in cybersecurity: Exploring the accidental insider. In: *Psychological and behavioral examinations in cyber security*. IGI Global; 2018. p. 46-63.
- [7] Fromme J, Biermann R, Unger A. Serious Games oder taking games seriously? *Digitale Lernwelten: Konzepte, Beispiele und Perspektiven*. 2010:39-57.
- [8] Jin G, Tu M, Kim TH, Heffron J, White J. Evaluation of game-based learning in cybersecurity education for high school students. *Journal of Education and Learning (EduLearn)*. 2018;12(1):150-8.
- [9] Khan MA, Merabet A, Alkaabi S, Sayed HE. Game-based learning platform to enhance cybersecurity education. *Education and Information Technologies*. 2022:1-25.
- [10] Mostafa M, Faragallah OS. Development of serious games for teaching information security courses. *IEEE Access*. 2019;7:169293-305.
- [11] Ejsing-Duun S. A game magically circling. In: *Serious games in education: a global perspective*. Aarhus Universitetsforlag; 2011. p. 170-86.
- [12] Nazri SNB, Kadir TABA, Ihsan SNB. Review on Serious-Education Games for Slow Learners. In: *2023 IEEE 8th International Conference On Software Engineering and Computer Systems (ICSECS)*. IEEE; 2023. p. 441-6.

- [13] Barth V, Wanner T, Gelderie M, Schüle J. Is attacking more fun than defending?—Observations from Developing and Evaluating Practical Game Based Learning. *Procedia Computer Science*. 2024;246:3148-57.
- [14] Stockhardt S, Reinheimer B, Volkamer M, Mayer P, Kunz A, Rack P, et al. Teaching phishing-security: which way is best? In: *ICT Systems Security and Privacy Protection: 31st IFIP TC 11 International Conference, SEC 2016, Ghent, Belgium, May 30-June 1, 2016, Proceedings 31*. Springer; 2016. p. 135-49.
- [15] Rikkers V, Sarmah DK. A story-driven gamified education on USB-based attack. *Journal of Computing in Higher Education*. 2025;37(1):248-72.
- [16] Kaplan-Rakowski R, Lin L, Wojdyski T. Learning vocabulary using 2D pictures is more effective than using immersive 3D stereoscopic pictures. *International Journal of Human-Computer Interaction*. 2022;38(4):299-308.
- [17] Zaharias P, Chatzeparaskevidou I, Karaoli F. Learning Geography Through Serious Games: The Effects of 2-Dimensional and 3-Dimensional. *Learning*. 2017;9(1).
- [18] Karavidas L, Lilis V, Tsiatsos T, Douka S. Choosing the Dimension of a Serious Game: The HALT Case Study. In: *2024 IEEE Global Engineering Education Conference (EDUCON)*. IEEE; 2024. p. 1-5.
- [19] Makransky G, Petersen GB. The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational psychology review*. 2021;33(3):937-58.
- [20] Aksoy E, et al. Comparing the effects on learning outcomes of tablet-based and virtual reality-based serious gaming modules for basic life support training: randomized trial. *JMIR serious games*. 2019;7(2):e13442.
- [21] Carbonell-Carrera C, Jaeger AJ, Saorín JL, Melián D, De la Torre-Cantero J. Minecraft as a block building approach for developing spatial skills. *Entertainment Computing*. 2021;38:100427.
- [22] Dan A, Reiner M, et al. Real time EEG based measurements of cognitive load indicates mental states during learning. *Journal of Educational Data Mining*. 2017;9(2):31-44.
- [23] Ak O, Kutlu B. Comparing 2D and 3D game-based learning environments in terms of learning gains and student perceptions. *British Journal of Educational Technology*. 2017;48(1):129-44.
- [24] Kaimara P, Fokides E, Oikonomou A, Atsikpasi P, Deliyannis I. Evaluating 2D and 3D serious games: The significance of student-player characteristics. *Διάλογοι! Θεωρία και πράξη στις επιστήμες αγωγής και εκπαίδευσης*. 2019;5:36-56.
- [25] Lane HC, Hays MJ, Auerbach D, Core MG. Investigating the relationship between presence and learning in a serious game. In: *Intelligent Tutoring Systems: 10th International Conference, ITS 2010, Pittsburgh, PA, USA, June 14-18, 2010, Proceedings, Part I 10*. Springer; 2010. p. 274-84.
- [26] Pixune. Game Art Outsourcing Price: How Much Does It Cost?; 2024. Accessed: 01-Apr-2025. Available from: <https://pixune.com/blog/game-art-outsourcing-price/>.
- [27] Awojana T, Chou TS. Overview of learning cybersecurity through game based systems. In: *2019 CIEC*; 2019. .
- [28] Angafor GN, Yevseyeva I, He Y. Game-based learning: A review of tabletop exercises for cybersecurity incident response training. *Security and privacy*. 2020;3(6):e126.
- [29] Karagiannis S, Magkos E. Engaging students in basic cybersecurity concepts using digital game-based learning: Computer games as virtual learning environments. *Advances in Core Computer Science-Based Technologies: Papers in Honor of Professor Nikolaos Alexandris*. 2021:55-81.
- [30] Bacud ML, Mases S. Game-based learning for cybersecurity awareness training programmes in the public sector. In: *ECEL 2021 20th European Conference on e-Learning. Academic Conferences International limited*; 2021. p. 50.
- [31] Roepke R, Schroeder U. Teaching defence against the dark arts using game-based learning: A review of learning games for cybersecurity education. In: *International Conference on Computer Supported Education*. Springer; 2019. p. 71-87.
- [32] Tobias S, Fletcher JD. *Computer Games and Instruction*. Charlotte, NC: Information Age Publishing; 2011.
- [33] Mayer RE. *Computer Games for Learning: An Evidence-Based Approach*. Cambridge, MA: MIT Press; 2014.
- [34] Irvine CE, Thompson MF, Allen K. CyberCIEGE: an information assurance teaching tool for training and awareness. *Naval Postgraduate School (US)*.; 2005.
- [35] Ashley TD, Kwon R, Gourisetti SNG, Katsis C, Bonebrake CA, Boyd PA. Gamification of cybersecurity for workforce development in critical infrastructure. *Ieee Access*. 2022;10:112487-501.
- [36] Cano J, Hernández R, Ros S, Tobarra L. A distributed laboratory architecture for game based learning in cybersecurity and critical infrastructures. In: *2016 13th International Conference on Remote Engineering and Virtual Instrumentation (REV)*. IEEE; 2016. p. 183-5.
- [37] Prümmer J, van Steen T, van den Berg B. Successful Gamification of Cybersecurity Training. *Cyberpsychology, Behavior, and Social Networking*. 2021;24(9):593-8.
- [38] Wu T, Tien KY, Hsu WC, Wen FH. Assessing the Effects of Gamification on Enhancing Information Security Awareness Knowledge. *Applied Sciences*. 2021;11(19):9266.
- [39] Parong J, Mayer RE. Learning science in immersive virtual reality. *Journal of Educational Psychology*. 2018;110(6):785-97.
- [40] Dede T Eren; Wanner, Gelderie M. Screenshot of 2D UI card game; 2025. Authors' original image; implemented with React.
- [41] Dede T Eren; Wanner, Gelderie M. Screenshot of 3D first-person hospital environment; 2025. Authors' original image; implemented in Godot engine.
- [42] Godot Engine contributors. Godot Engine;. Open-source game engine; accessed 14 August 2026. <https://godotengine.org/>.
- [43] Meta Platforms, Inc . React – A JavaScript library for building user interfaces; 2026. Accessed: 24 August 2025. <https://reactjs.org/>.

- [44] Bloom BS, Engelhart MD, Furst EJ, Hill WH, Krathwohl DR. Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain. Longmans, Green; 1956.
- [45] Anderson LW, Krathwohl DRe. A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. Longman; 2001.
- [46] Bundesamt für Sicherheit in der Informationstechnik (BSI). Pflichten für KRITIS-Betreiber; 2025. Accessed: 1 March 2025. <https://www.bsi.bund.de/dok/12211858>.
- [47] Bundesministerium des Innern (BMI). Nationale Strategie zum Schutz Kritischer Infrastrukturen (KRITIS-Strategie); -. Accessed: 16 March 2025. https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/bevoelkerungsschutz/kritis.pdf?__blob=publicationFile&v=3.
- [48] OpenKRITIS. Das NIS2-Umsetzungsgesetz (NIS2UmsuCG); 2025. Accessed: 3 March 2025. <https://www.openkritis.de/it-sicherheitsgesetz/nis2-umsetzung-gesetz-cybersicherheit.html>.
- [49] Prolific Academic Ltd .Prolific; 2025. Accessed: 28-Apr-2025. Available from: <http://www.prolific.com>.
- [50] Tivian. Unipark | Tivian DE; 2025. Accessed: 15-Apr-2025. Available from: <http://www.tivian.com>.
- [51] Fu FL, Su RC, Yu SC. EGameFlow: A scale to measure learners' enjoyment of e-learning games. Computers & Education. 2009;52(1):101-12.
- [52] Welch BL. The generalisation of "Student's" problem when several different population variances are involved. Biometrika. 1947;34(1-2):28-35.
- [53] Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
- [54] Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika. 1965;52:591-611.
- [55] Virtanen P, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods. 2020;17:261-72.
- [56] Python Software Foundation. Python Language Reference; 2023. Version 3.12, <https://www.python.org/>.
- [57] Levene H. Robust tests for equality of variances. Contributions to Probability and Statistics. 1960:278-92.